

Université de Sherbrooke

Interactions et actions de hnRNP A1/UP1 et caractérisation
des extensions télomériques simple-brin chez la levure et chez l'humain.

Par

Catherine LeBel

Thèse présentée à la Faculté de Médecine

En vue de l'obtention du grade de

Philosophae doctor (Ph.D.) en microbiologie

Avril 2005



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-14863-1

Our file Notre référence

ISBN: 978-0-494-14863-1

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

TABLE DES MATIÈRES

Table des matières.....	I
Liste des figures.....	IV
Liste des abréviations.....	VII
Historique des études graduées et des projets.....	XI
Résumé.....	XV
Introduction.....	1
Les télomères.....	1
La structure générale d'un télomère : ADN et protéines.....	4
L'ADN télomérique.....	4
Les différentes fonctions associées aux télomères.....	9
Les différentes fonctions associées aux télomères : la protection des chromosomes.....	9
Les différentes fonctions associées aux télomères : la réplication complète des chromosomes.....	17
L'implication des télomères dans la détermination du futur des cellules..	22
La structure 3' terminale des extrémités des chromosomes et son importance.....	25
Objectifs du projet.....	37
Généralités concernant l'épissage et l'épissage alternatif.....	39
Les mécanismes et la machinerie de base de l'épissage.....	42
L'épissage alternatif.....	50
Les protéines hnRNP A1 et hnRNP A2.....	56

Objectifs du projet.....	64
Chapitre 1.....	66
Préambule.....	66
Matériel et méthodes.....	67
Résultats.....	80
Chapitre 2.....	111
Préambule 1.....	111
Préambule 2	145
Chapitre 3.....	147
Préambule.....	147
Chapitre 4.....	181
Préambule.....	181
Matériel et méthodes.....	183
Résultats.....	187
Discussion.....	201
Interactions entre les molécules hnRNP A1 et implications possibles aux télomères.....	201
Interactions de hnRNP A1 et modèle d'action dans l'épissage alternatif.....	207
Rôles de hnRNP A2 dans la modulation de l'épissage alternatif.....	218
Régulation de l'expression de A1 et A1 ^B : de nombreux éléments de contrôle..	220
Existence des extensions télomériques en phase G1 du cycle cellulaire chez la levure.....	222
Le complexe MRX serait l'exonucléase 5'-3' responsable de la génération et du	

maintien des extensions 3' simple-brin chez la levure.....	231
Analyse des extensions télomériques simple-brin de lignées de cellules humaines avec une méthode de T-OLA modifiée.....	241
Maintien des télomères dans des cellules ALT : modèles possibles.....	247
Conclusions.....	251
Remerciements.....	252
Références.....	254
Annexe	332
Préambule.....	332

LISTE DES FIGURES

Figure 1. Structure générale d'un télomère.....	4
Figure 2. Structure d'un télomère de la levure <i>S. cerevisiae</i>	7
Figure 3. Structure d'un télomère humain.....	8
Figure 4. Structure des t-loop et D-loop.....	15
Figure 5. Modèle pour la formation de l'hétérochromatine télomérique chez <i>S. cerevisiae</i>	17
Figure 6. Schéma de la réplication de l'ADN et du problème du brin avancé lors de la réplication des télomères (end-replication problem).....	18
Figure 7. Modèle de l'élongation des télomères par la télomérase.....	19
Figure 8. Changements dans la protection des télomères.....	25
Figure 9. Multiples signaux reconnus par les facteurs d'épissage lors de la formation du spliceosome.....	43
Figure 10. Réactions de trans-estérification menant à l'épissage des ARNs pré- messagers.....	44
Figure 11. Formation du spliceosome et des différents complexes sur l'intron.....	46
Figure 12. Facteurs et sites d'épissage permettant la définition de l'exon.....	48
Figure 13. Régulation de l'épissage chez la drosophile.....	51
Figure 14. Régulation de l'épissage de l'élément P chez la drosophile.....	52
Figure 15. Différents patrons possibles pour l'épissage alternatif.....	54
Figure 16. Famille des protéines hnRNP A1/A2.....	56
Figure 17. Éléments introniques conservés de la protéine hnRNP A1.....	59
Figure 18. Modèle d'action de hnRNP A1.....	60

Figure 19. Double-hybride chez la levure et plasmides utilisés.....	67
Figure 20. La protéine GAL4 DBD-UP1 peut lier un oligo télomérique humain.....	83
Figure 21. Caractérisation des interactions entre les domaines de la protéine A1 par l'utilisation du système double-hybride.....	93
Figure 22. Séquence des acides aminés de UP1 et structure de la protéine.....	94
Figure 23. Aperçu de la recombinaison homologe menant à un plasmide linéaire.....	98
Figure 24. Analyses de la perte d'interaction entre les protéines hnRNP A1 contenant des mutations aléatoires.....	99
Figure 25. Gels de retardement des protéines recombinantes GST-A1 mutantes avec un oligo ARN.....	102
Figure 26. Épissage <i>in vitro</i> avec les protéines GST-A1 mutantes.....	107
Figure 27. Schéma des étapes de la réaction de T-OLA.....	188
Figure 28. Réactions de T-OLA modifiées; différentes conditions de réaction et différents substrats.....	192
Figure 29. Les réactions de T-OLA ont lieu sur la portion 3' simple-brin des extrémités télomériques de l'ADN purifié de cellules HeLa S3.....	194
Figure 30. Réactions de T-OLA sur l'ADN purifié de différentes lignées cellulaires (1).....	196
Figure 31. Réactions de T-OLA sur l'ADN purifié de différentes lignées cellulaires (2).....	200
Figure 32. Modèle d'action des protéines hnRNPs aux télomères.....	206
Figure 33. Modèle pour la réplication des télomères de levure et pour l'établissement de la structure télomérique observée en phase G1.....	240

Figure 34. Mécanismes possibles pour le maintien des télomères dans les cellules

ALT.....	250
Tableau 1. Séquence des répétitions télomériques de différents organismes.....	4
Tableau 2. Caractéristiques des répétitions télomériques de différents organismes.....	31
Tableau 3. Caractéristiques des souches de levure et lignées cellulaires utilisées.....	79
Tableau 4. Synthèse des protéines pouvant interagir avec la protéine UP1.....	86
Tableau 5. Résumé des résultats obtenus avec les protéines hnRNP A1 contenant des mutations à des résidus importants pour certaines interactions des protéines de type sauvage.....	95
Tableau 6. Résumé des données récupérées suite aux analyses faites sur les protéines de fusion GST-A1 contenant des mutations aléatoires.....	101
Tableau 7. Caractéristiques des lignées cellulaires étudiées.....	184

LISTE DES ABRÉVIATIONS

AD	“activation domain”; domaine d’activation
ALT	“alternative lengthening of telomeres”; allongement alternatif des télomères
APB	“ALT-associated PML bodies”; corps PML associés au processus ALT
A2RE	“hnRNP A2 response element”; élément de réponse de hnRNP A2
ARS	“autonomously replicating sequence”; séquence ayant une réplication autonome
3-AT	3-aminotriazole
ATP	adénosine 5’-triphosphate
BSA	“bovine serum albumin”; albumine de serum bovin
CE	“conserved element”; element conservé
CFP	“circular form of the plasmid”; forme circulaire du plasmide
chIP	“chromatin immunoprecipitation”; immunoprécipitation de la chromatine
Ci	Curie
DBD	“DNA binding domain”; domaine de liaison à l’ADN
D-loop	“displacement loop”; boucle de déplacement
DSCAM	“Down syndrome cell adhesion molecule”; molécule d’adhésion cellulaire du syndrome Down
Dsx	“doublesex”; double-sexe
DTT	dithiothréitol
e.g.	“exempli gratia”; par exemple
ESE	“exonic splicing enhancer”; stimulateur exonique d’épissage

ESS	“exonic splicing silencer”; inhibiteur intronique d’épissage
EST	“ever shorter telomere”; télomères toujours plus courts
EST	“expressed sequence tags”; séquence exprimée étiquetée
FISH	“fluorescence <i>in situ</i> hybridization”; hybridation avec fluorescence <i>in situ</i>
GAL 4-AD	“GAL4 activation domain”; domaine d’activation de GAL4
GAL 4-DBD	“GAL4 DNA binding domain”; domaine de liaison à l’ADN de GAL4
GRD	“glycin-rich domain”; domaine riche en glycines
hnRNA	“heterogeneous nuclear ribonucleic acid”; acide ribonucléique nucléaire hétérogène
hnRNP	“heterogeneous nuclear ribonucleoparticle”; ribonucléoparticule nucléaire hétérogène
HPLC	“high pressure liquid chromatography”; chromatographie liquide sous haute pression
i.e.	“id est”; c’est-à-dire
ISE	“intronic splicing enhancer”; stimulateur intronique d’épissage
ISS	“intronic splicing silencer”; inhibiteur intronique d’épissage
Kb	kilobases
NHEJ	“nonhomologous end-joining”; ligation des extrémités non-homologues
Nt	nucléotides
OB fold	“oligonucleotide/oligosaccharide binding fold”; motif de liaison oligonucléotide/oligosaccharide
Pb	paire de bases
PCR	“polymerase chain reaction”; réaction de polymérisation en chaîne

PD	“population doubling”; doublement de population
PENT	“primer extension/nick translation”; extension d’amorce/remplissage de brèche
PML	“promyelocytic leukemia”; leucémie promyélocytique
PNA	“peptid nucleic acid”; peptide d’acides nucléiques
PSI	“P element somatic inhibitor” ; inhibiteur somatique de l’élément P
RNAi	“RNA interference”; interférence par l’ARN
RNP	“ribonucleoprotein”; ribonucléoprotéine
RRM	“RNA Recognition Motif”; motif de reconnaissance de l’ARN
RS	sérine/arginine
SDS	sodium dodecyl sulfate
SF1/BBP	“Splicing factor 1/ branchpoint binding protein”; facteur d’épissage 1 /protéine de liaison du point de branchement
snRNA	“small nuclear RNA”; petit ARN nucléaire
snRNP	“small nuclear ribonucleoparticle”; petite ribonucléoparticule nucléaire
SR	“serine/arginine”; serine/arginine
STELA	“single telomere length analysis”; analyse de longueur d’un télomère unique
SV40	“simian virus 40”; virus simien 40
Sxl	“sex-lethal”; sexe léthal
TAS	“telomere associated sequences”; sequences associées aux télomères
TEBP	“telomere end binding protein”; protéine de liaison des extrémités télomériques

t-loop	“telomere loop”; boucle télomérique
T-OLA	“telomere oligonucleotide ligation assay”; essai de liagtion d’oligonucléotides télomériques
T4 PNK	“T4 polynucleotide kinase”; polynucléotide kinase du phage T4
Tra	“transformer”; transformeur
TRAP	“telomeric repeat amplification protocol”; protocole d’amplification de répétitions télomériques
TRD	“telomere rapid deletion”; délétion rapide de télomères
TRF	“terminal restriction fragment”; fragment de restriction terminal
TRF2	“telomere binding repeat factor 2”; facteur de liaison des répétitions télomériques 2
U2AF	“U2 snRNP auxiliary factor”; facteur auxiliaire U2
UTR	“untranslated region”; région non-traduite
VIH-1	virus d’immunodéficience humaine de type 1
V/V	volume/volume
W/V	“weight/volume”; poids/volume

HISTORIQUE DES ÉTUDES GRADUÉES ET DES PROJETS

Pendant l'été 1996, j'ai eu le plaisir d'être initiée à la biologie moléculaire comme étudiante d'été dans le laboratoire du Dr. Benoit Chabot. Ce stage d'été m'a grandement aidé dans ma décision de poursuivre des études graduées dans ce domaine. En septembre 1997, j'ai commencé une maîtrise en co-direction au département de microbiologie, dans deux laboratoires ayant des sujets de recherche passablement différents; le laboratoire du Dr. Benoit Chabot étudiait la modulation de l'épissage alternatif chez les mammifères, tandis que le laboratoire du Dr. Raymund J. Wellinger étudiait la réplication des télomères chez la levure *Saccharomyces cerevisiae*. Une collaboration toute spéciale entre ces deux laboratoires venait de montrer que la protéine hnRNP A1/UP1 (aussi appelée A1 et connue pour son activité dans la modulation de l'épissage alternatif) avait un rôle important à jouer dans la biogénèse des télomères (LaBranche et al. 1998).

Suite à la publication de cet article, la protéine hnRNP A1/UP1 est devenue le lien entre 2 sujets qui semblaient au départ diamétralement opposés, et a donné naissance à mon premier projet de maîtrise. Ce projet était de trouver par un criblage double-hybride chez la levure, des partenaires cellulaires de hnRNP A1/UP1 qui pourraient être impliqués avec A1 dans l'un ou l'autre des processus cellulaires décrits auparavant. Plusieurs autres projets ont d'ailleurs suivi sur ces sujets dans le laboratoire du Dr. Benoit Chabot, donnant lieu à plusieurs publications (Blanchette et Chabot 1999; Côté et Chabot 1999; Dallaire et al. 2000; Simard et Chabot 2000; Fiset et Chabot 2001; Bolduc et al. 2001; Hutchison et al. 2002; Simard et Chabot 2002; Nasim et al. 2002; Patry et al. 2003; Patry et al. 2004).

En plus de mon projet de criblage, dès le début de ma maîtrise, j'ai travaillé sur un autre projet qui visait l'étude des télomères chez l'organisme modèle *S. cerevisiae*. Ce projet était de caractériser l'extension télomérique 3' simple-brin, en mettant au point une méthode de mesure de cette extension chez la levure. Ce projet était basé sur un article publié par le laboratoire du Dr. Raymund J. Wellinger qui décrivait une méthode pour la mesure des extensions 3' simple-brin, chez les cellules humaines (McElligott et Wellinger 1997). Cependant, les différences notables entre les télomères des deux espèces ont fait que de nombreuses approches biochimiques ont été essayées (extension d'amorce comme chez les mammifères, clivage au KMnO₄, clivage au DEPC), avant d'en arriver à en mettre une au point.

En 1999, le travail de Marco Blanchette dans le laboratoire du Dr. Benoit Chabot a mené à l'élaboration d'un modèle de « looping-out » pour expliquer le rôle de hnRNP A1 dans la sélection des sites d'épissage alternatif. Dans ce modèle, il doit y avoir une interaction de la protéine hnRNP A1 avec elle-même pour permettre le « looping out ». Avec ces données, j'ai utilisé un système de mutagénèse aléatoire par PCR sur la protéine hnRNP A1 et le système de double-hybride afin de mieux caractériser l'interaction de hnRNP A1 avec elle-même. Ce projet m'a permis de me familiariser avec un autre aspect de la biologie moléculaire, dont le travail avec des extraits nucléaires de mammifères, la production et la purification des protéines recombinantes, la transcription *in vitro* et les réactions d'épissage *in vitro*. De plus, le système de double-hybride que j'utilisais a été utile pour confirmer certaines données d'interactions *in vitro* obtenues par d'autres étudiants. Le projet de caractérisation de la protéine A1 a été poursuivi par Benoit Labrecque dans le laboratoire du Dr. Chabot et un projet semblable a aussi été commencé

par Bruno Lemieux chez Télégène Inc., et j'ai été impliquée à différentes étapes au cours de ces différents projets.

Un projet concernant la protéine hnRNP A2 en collaboration avec le groupe du Dr. Andrew Mouland de l'Université de Montréal a été amorcé dans le laboratoire. Le projet visait à vérifier si deux protéines de la même famille (hnRNP A1 et A2) pouvaient avoir des modes d'action similaires dans la sélection des sites d'épissage alternatif. Ce projet m'a permis de travailler en collaboration avec Stephen Hutchison et a mené à deux publications (Hutchison et al. 2002; Bériault et al. 2004), où mes implications ont été diverses.

En avril 2001, au retour d'un congrès, Dr. Raymund J. Wellinger me propose de reprendre mes études sur l'extension télomérique simple-brin de la levure, en adaptant un système utilisé auparavant chez *Tetrahymena* (Jacob et al. 2001). Encore une fois, les particularités des télomères de la levure font en sorte que ce système, bien qu'il donne certains résultats encourageants, ne permet pas d'atteindre notre but. Par la suite, nous avons tenté en vain de mettre au point une approche de PCR sur les télomères et finalement, ce fut une approche d'extension d'amorce modifiée qui nous a permis de trouver une façon de mesurer le simple-brin télomérique chez la levure (Larrivée et al. 2004).

Dernièrement, nous avons décidé d'étudier les extensions télomériques chez les cellules humaines, en raffinant une méthode mise au point par un autre laboratoire (Cimino-Reale et al. 2001). Les modifications apportées à la technique ont permis de collaborer avec Télégène Inc., et malgré le fait que les résultats obtenus jusqu'à

maintenant soient encourageants, il reste encore du travail à effectuer sur cette portion de projet.

Le déroulement de mes études graduées permet de voir qu'avec le temps et les résultats obtenus, les projets sur lesquels je travaillais au départ sont devenus de plus en plus difficiles à relier entre eux. Selon les différentes périodes, j'ai pu travailler avec la levure et/ou avec les cellules de mammifères, sur l'épissage alternatif et/ou les télomères. La thèse sera donc un résumé des différents travaux que j'ai accomplis sur ces deux sujets qui au départ étaient étroitement reliés, mais qui au fil des ans se sont séparés.

RÉSUMÉ

Les protéines hnRNP A1/UP1 et hnRNP A2 sont des protéines nucléaires abondantes qui possèdent plusieurs rôles dans la cellule, en particulier dans l'épissage alternatif et la biogénèse des télomères. L'action de ces protéines dans l'épissage alternatif semble directe, tandis que leur action au niveau des télomères pourrait être indirecte. Pour mieux caractériser le rôle de hnRNP A1/UP1 aux télomères, nous avons utilisé une approche de criblage double-hybride dans la levure afin d'identifier leurs partenaires cellulaires. Avec ce criblage, nous avons récupéré plusieurs protéines différentes, mais aucune protéine ayant une implication ou une localisation aux télomères n'a été récupérée. Cependant, une forte interaction entre des molécules hnRNP A1 elles-mêmes a été montrée avec ce criblage. Une telle interaction entre des protéines hnRNP A1 est à la base d'un modèle qui a été proposé pour expliquer le rôle de hnRNP A1 dans la modulation de l'épissage alternatif. En utilisant diverses approches de mutagenèse, nous avons tenté de déterminer les domaines de hnRNP A1 requis pour cette modulation. Les résultats obtenus suggèrent que le modèle élaboré pour l'explication du mécanisme d'action de hnRNP A1 dans l'épissage est valide, c'est-à-dire qu'un mécanisme comme celui qui est proposé, pourrait servir à moduler l'épissage alternatif de hnRNP A1. De plus, nous avons montré que la protéine hnRNP A2, qui fait partie de la même famille de protéines que hnRNP A1, possède des fonctions similaires à celles de hnRNP A1 dans la modulation de l'épissage alternatif. En utilisant deux différents groupes de transcrits modèles, nous avons montré que hnRNP A2 peut se substituer à hnRNP A1 dans certaines conditions d'épissage, mais ce processus est dépendant de la présence des sites de liaison optimaux dans l'ARN.

Les télomères sont les extrémités des chromosomes eucaryotes. Ils sont constitués de répétitions de courtes séquences, dont un des brins est riche en guanines (G). Chez toutes les espèces étudiées, ce brin se retrouve à l'extrémité des chromosomes sous forme d'une extension 3' simple-brin. Lors de la réplication des télomères chez la levure *Saccharomyces cerevisiae*, les extrémités des chromosomes acquièrent de longues extensions 3' simple-brin riches en G (plus de 30 bases). La présence de ces extensions en dehors de la phase S (phase où a lieu la réplication de l'ADN) est inconnue chez la levure. Également, les protéines impliquées dans la génération et le maintien de ces extensions sont peu connues. Avec la mise au point d'une nouvelle approche d'extension d'amorce modifiée, mes travaux ont permis la caractérisation de l'extension télomérique 3' simple-brin chez la levure en phase G1 du cycle cellulaire (12-14 bases). De plus, nos résultats ont montré que la vaste majorité des télomères possèdent une extension 3' simple-brin en dehors de la phase S. Nous avons aussi montré l'implication d'un complexe protéique contenant une activité exonucléase (MRX) dans la génération et le maintien de ces extensions. Comme l'extension 3' simple-brin est une caractéristique essentielle pour la protection des chromosomes, des activités exonucléases redondantes à MRX seraient aussi présentes dans les cellules afin d'assurer la conformation adéquate des télomères. Chez les cellules humaines, la présence de cette extension simple-brin est critique afin d'assurer la protection des extrémités des chromosomes. Nous avons modifié une approche utilisée pour la mesure du simple-brin chez les cellules humaines, afin d'augmenter la précision des résultats obtenus dans la mesure des extensions simple-brin. Cette approche modifiée a d'abord été validée sur des oligonucléotides, puis appliquée à

l'ADN génomique pour confirmer les mesures des extensions simple-brin obtenues auparavant. Par la suite, nous avons commencé la caractérisation des extensions simple-brin dans des cellules transformées exprimant la télomérase et dans des cellules transformées qui maintiennent leurs télomères par un processus alternatif à la télomérase (ALT). Les résultats obtenus par notre approche modifiée pour la mesure de l'extension simple-brin présente dans les cellules transformées (plus de 200 bases) confirment ceux qui ont été obtenus auparavant par d'autres groupes (Cimino-Reale et al. 2001; Stewart et al. 2003). Les résultats également obtenus par cette approche modifiée, montrent que la taille des extensions télomériques présentes dans des cellules ALT serait plus courte que dans les autres types de cellules, ce qui constitue une première donnée à ce sujet.

INTRODUCTION

Les télomères

Suite à l'évolution des espèces, les génomes eucaryotes se sont différenciés des génomes procaryotes par l'acquisition ou le développement de chromosomes linéaires plutôt que circulaires. Cette nouvelle linéarité des chromosomes implique, entre autres, l'apparition d'extrémités chromosomiques. Par conséquent, tous les organismes qui possèdent des chromosomes linéaires ont besoin de différencier ceux-ci des cassures double-brin de l'ADN, et de protéger ces extrémités chromosomiques contre des activités de dégradation, de réparation ou de ligation. En effet, si ces différentes activités étaient présentes aux extrémités des chromosomes, elles mettraient en péril la stabilité du génome. Les premières études visant à étudier ce phénomène remontent à 1938, et furent menées par le généticien H. J. Muller, qui procéda à des irradiations aux rayons X des chromosomes de drosophile. Il observa alors que des délétions ou des inversions dans les régions terminales des chromosomes étaient quasiment indétectables, alors qu'elles se trouvaient facilement dans les autres régions des chromosomes (Muller 1938). Sa conclusion fut que les extrémités des chromosomes, qu'il nomma télomères, possèdent une structure particulière qui leur permet de stabiliser les chromosomes.

Environ pendant la même période, les études de B. McClintock sur le maïs ont permis de montrer que des chromosomes ayant subi des cassures dans l'ADN double-brin sont capables de fusionner entre eux, tandis que les extrémités naturelles des chromosomes ne fusionnent pas entre elles (McClintock 1939; McClintock 1941). Ces résultats suggèrent que les extrémités des chromosomes possèdent des caractéristiques qui les diffèrent des cassures d'ADN double-brin. Depuis ce temps, des analyses

approfondies de la structure et de la composition des télomères ont permis de mettre en évidence que ceux-ci sont constitués par l'association d'ADN télomérique et de protéines spécifiques, pour former ce que nous appelons le capuchon télomérique. Les télomères sont donc des complexes spécialisés qui sont présents aux extrémités des chromosomes eucaryotes, et qui possèdent plusieurs fonctions essentielles (voir plus loin).

Plusieurs organismes servent de modèles pour l'étude de la biologie des télomères. Ces organismes sont les protozoaires ciliés (*Tetrahymena thermophila*, *Tetrahymena paravorax*, *Oxytricha nova*, *Oxytricha fallax*, *Euplotes crassus*, *Euplotes aediculatus*), les levures de fission (*Schizosaccharomyces pombe*) ou de bourgeonnement (*Saccharomyces cerevisiae*, *Kluveromyces lactis*), les cellules de mammifères (lignées de cellules humaines ou murines), la mouche (*Drosophila melanogaster*) et les plantes (*Arabidopsis thaliana* et *Silene latifolia*). Chacun de ces modèles est important dans le domaine de recherche sur les télomères, et chacun possède des avantages qui ont mené à une meilleure compréhension de la biologie des télomères. D'abord, à cause de l'organisation particulière de leur génome et de l'abondance des télomères et de la télomérase, les ciliés sont des organismes modèles très utilisés. En effet, ces organismes possèdent deux noyaux; le micronoyau est diploïde et transcriptionnellement inactif et sert uniquement à la transmission de l'information génétique au cours de la reproduction, et le macronoyau, qui est issu du micronoyau lors de la reproduction, qui est constitué de chromosomes fragmentés qui sont transcriptionnellement actifs (Jahn et Klobutcher 2002). Lors du processus de reproduction, les chromosomes sont fragmentés, et chaque petit fragment produit acquiert des télomères au site de coupure (Coyne et al. 1996). La grande quantité de télomères ainsi produite a permis la caractérisation des télomères et

des extensions télomériques (Klobutcher et al. 1981), et a mené à la découverte de l'enzyme responsable du maintien des télomères (Greider et Blackburn 1985) (voir plus loin). Les levures sont également des organismes qui sont très utilisés pour l'étude des télomères en raison de leur génétique simple et puissante, de la diversité des outils de biologie moléculaire qui peuvent être utilisés, et de la relative conservation observée entre la levure et l'humain au niveau de la structure des télomères et des facteurs qui leur sont associés. En fait, la majorité des connaissances de base sur la biologie des télomères vient des études qui ont été faites chez la levure *S. cerevisiae*. Les études avec *S. pombe* ont permis de montrer des similarités étonnantes entre la levure de fission et les cellules humaines, et les études avec *K. lactis* ont permis de mieux comprendre certains procédés alternatifs reliés au maintien des télomères et la caractérisation de l'ARN de la télomérase. L'utilisation de la drosophile a permis d'étudier certaines différences possibles au niveau des télomères, car la mouche et certains autres insectes possèdent aux extrémités de leurs chromosomes de multiples copies de rétrotransposons spécifiques plutôt que des séquences télomériques répétées. L'utilisation de lignées de cellules de mammifères permet l'étude de processus cellulaires importants reliés aux télomères et aux protéines télomériques, et ce malgré des différences énormes qui existent entre différentes lignées de cellules qui peuvent être utilisées. Encore de nos jours, les organismes les plus utilisés pour la recherche sur les télomères sont les ciliés, les levures et les cellules de mammifères, et seront discutés plus en détails dans le texte. Ces organismes ont permis de faire d'importantes découvertes dans le vaste domaine de recherche sur les télomères, et restent toujours des modèles utiles pour approfondir nos connaissances de la biologie des télomères.

La structure générale d'un télomère : ADN et protéines

Chez tous les organismes, les télomères sont composés d'ADN et de protéines spécifiques qui s'y lient, et sont situés aux extrémités des chromosomes. Les télomères sont constitués d'une portion d'ADN télomérique double-brin, qui est de taille variable selon les organismes, et d'une extension terminale d'ADN télomérique simple-brin, qui varie aussi en taille selon les organismes (figure 1). Chez la plupart des organismes, des régions sous-télomériques sont également présentes et composées de séquences répétées appelées TAS (telomere-associated sequences), qui sont elles aussi composées d'éléments répétés (Henderson 1995). Ces séquences semblent très dynamiques et peuvent varier grandement en taille entre les espèces et au sein d'une même espèce.

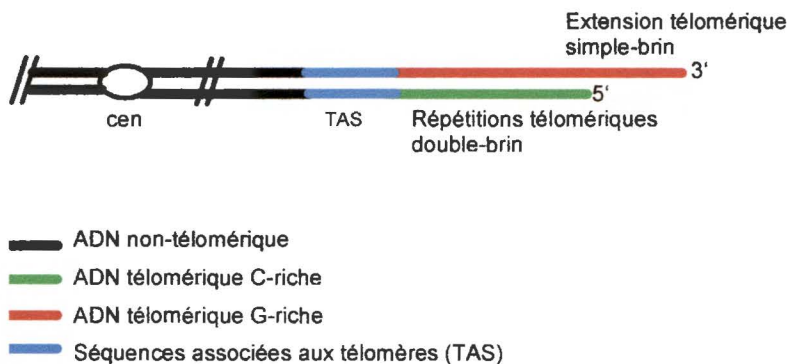


Figure 1. Structure générale d'un télomère. En partant du centromère vers la fin du chromosome, il y a d'abord l'ADN non-télomérique (noir), les séquences associées aux télomères (bleu), les répétitions télomériques double-brin (un brin d'ADN en rouge et l'autre brin d'ADN en vert) et finalement l'extension télomérique simple-brin (rouge). Des protéines spécifiques peuvent se lier à toutes les séquences d'ADN décrites.

L'ADN télomérique

La première séquence télomérique à être déterminée fut celle du protozoaire cilié *Tetrahymena thermophila* (Blackburn et Gall 1978). L'isolation et le séquençage de l'ADN télomérique issu de cette espèce furent facilités par la présence d'un vaste nombre de télomères dans le macronoyau transcriptionnellement actif. L'ADN télomérique de cet organisme est constitué de courtes séquences répétées en tandem. Les télomères de cet

organisme comportent de 50 à 70 répétitions de la séquence TTGGGG/CCCCAA (T₂G₄/C₄A₂) (Blackburn et Gall 1978).

Aujourd'hui, le séquençage de l'ADN télomérique chez différentes espèces montre que certaines caractéristiques de cet ADN sont relativement bien conservées. Dans la plupart des cas, l'ADN télomérique est constitué de séquences répétées en tandem. Un brin contient des répétitions toujours riches en résidus guanine (G) et ce brin G-riche est toujours orienté de 5' à 3' vers l'extrémité du chromosome (Tableau 1) (revue dans (Wellinger et Sen 1997) et LeBel, C. et Wellinger, R.J. (2005) J.Cell Sci., sous presse).

Organismes	Séquences répétées	Références
Protozoaires		
<i>Tetrahymena</i>	T ₂ G ₄	(Blackburn et Gall 1978)
<i>Oxytricha</i>	T ₄ G ₄	(Klobutcher et al. 1981)
<i>Euplotes</i>	T ₄ G ₄	(Klobutcher et al. 1981)
<i>Trypanosoma</i>	T ₂ AG ₃	(Blackburn et Challoner 1984) (Van der Ploeg et al. 1984)
Champignons		
<i>Saccharomyces</i>	(TG) ₁₋₆ TG ₂₋₃	(Shampay et al. 1984)
<i>Schyzosaccharomyces</i>	T ₁₋₂ ACA ₀₋₁ C ₀₋₁ G ₁₋₆	(Sugawara et Szostak 1986)
Plantes		
<i>Arabidopsis</i>	T ₃ AG ₃	(Richards et Ausubel 1988)
Vertébrés		
<i>Homo Sapiens</i>	T ₂ AG ₃	(Moyzis et al. 1988)
Plusieurs autres vertébrés	T ₂ AG ₃	(Meyne et al. 1989)

Tableau 1: Séquence des répétitions télomériques de différents organismes. La séquence présentée est celle du brin G-riche et va de 5' à 3' vers l'extrémité du chromosome.

Chez tous les organismes étudiés, le brin G-riche est plus long que le brin complémentaire riche en cytosines (C-riche), et forme ainsi une extension simple-brin en 3' du chromosome. Cependant, la séquence des répétitions télomériques varie selon les espèces, et le nombre de ces répétitions télomériques varie selon les espèces, selon les individus et même entre les différents chromosomes d'une même cellule. De nombreuses

études ont permis de dévoiler la taille des télomères chez différentes espèces. Cette longueur des télomères varie grandement entre les espèces étudiées, allant de 38 pb dans le macronoyau des ciliés *Oxytricha* et *Euplotes*, jusqu'à plusieurs dizaines de kilobases chez les vertébrés.

La levure constitue un excellent organisme modèle pour l'étude de la biologie des télomères. Elle possède une génétique qui est relativement simple, son génome est séquencé depuis plusieurs années, et de multiples banques de données sont disponibles sur la levure. De plus, la levure est un organisme modèle grandement étudié et utilisé dans divers domaines, et nous sert d'organisme modèle pour nos études sur les télomères. Chez la levure *S. cerevisiae*, les télomères comportent une séquence riche en G dégénérée, qui est $(TG)_{1-6}TG_{2-3}$ (que l'on abrège TG_{1-3}), et chacun des télomères compte environ 300 ± 75 pb de ces séquences télomériques (Shampay et al. 1984; Wang et Zakian 1990). Les télomères comportent aussi une extension télomérique 3' simple-brin de taille variable selon les phases du cycle cellulaire (voir plus loin). Les chromosomes de la levure comportent aussi des régions sous-télomériques très dynamiques et variables en taille. Deux classes de séquences (ou éléments) associées aux télomères peuvent être présentes. Tout d'abord, il y a l'élément conservé Y', qui est présent sur environ 70% des télomères de levure. Cet élément peut être de deux tailles possibles soit 6.7 kb (appelé Y' long) ou 5.2 kb (appelé Y' court) (Chan et Tye 1983; Walmsley et al. 1984; Louis et Haber 1992). Sur les 70% des télomères qui possèdent l'élément Y', celui-ci peut être présent de une à quatre fois. Dans cet élément Y', il y a présence de sites uniques pour des enzymes de restriction. Lorsque l'ADN génomique extrait de levures est digéré par certains de ces enzymes, il y a libération d'un fragment de restriction terminal appelé

TRF (Terminal Restriction Fragment) qui est caractéristique des souches de levure (Louis 1995). L'autre élément présent aux télomères de levure est l'élément X. Cet élément contient une origine de réplication qui est présente sur 31 des 32 télomères de levure (Louis et al. 1994). L'élément X est généralement placé plus près du centromère que l'élément Y' (figure 2). Étant donné que la présence ou l'absence de ces éléments aux télomères n'a pas de conséquences pour la levure, ces régions ne semblent pas avoir de fonctions essentielles (Wellinger et Zakian 1989). Par contre, ces régions peuvent avoir des rôles dans le maintien de l'architecture nucléaire, la liaison de facteurs spécifiques, l'établissement de la chromatine télomérique et les processus de recombinaison homologue au niveau des télomères (Pryde et Louis 1997; Pryde et Louis 1999; Teng et al. 2000; Grunstein 1997).

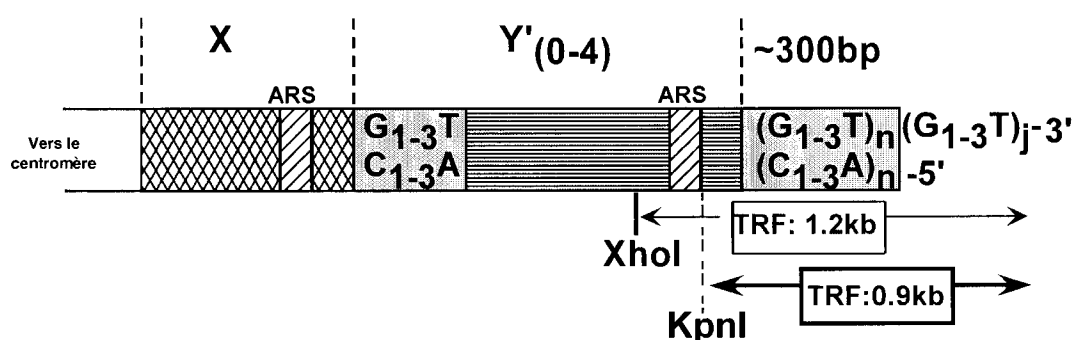


Figure 2. Structure d'un télomère de la levure *S. cerevisiae*. Les télomères présents aux extrémités des chromosomes possèdent 300 ± 75 pb de répétitions $C_{1-3}A/TG_{1-3}$. Les télomères possèdent aussi des éléments sous-télomériques X et Y' (description dans le texte).

Chez la plupart des vertébrés, incluant les mammifères et l'humain, les répétitions télomériques riches en G sont régulières, et comportent toujours des répétitions des 6 nucléotides T_2AG_3 (Moyzis 1991). Ces répétitions sont très variables en longueur, allant de 2 kb dans certains cas de lignées cellulaires transformées jusqu'à plus de 100 kb chez la souris (de Lange et al. 1990; Kipling et Cooke 1990; Starling et al. 1990). Les régions

sous-téломériques des télomères de mammifères sont encore mal identifiées et/ou caractérisées. Cependant, une région appelée X d'environ 2 à 4 kb a été identifiée, et semble représenter un segment d'ADN sous-téломérique qui est résistant aux digestions par des enzymes de restriction (Levy et al. 1998). Une étude montre également que des sites de l'ADN sous-téломérique peuvent être résistants à des digestions enzymatiques, selon les changements dans la longueur des télomères (Steinert et al. 2004). Les régions sous-téломériques chez les mammifères varient donc grandement. Par conséquent, aucun site unique pour un enzyme de restriction n'est présent de façon équidistante dans la région de transition entre l'ADN téломérique et l'ADN sous-téломérique chez les mammifères, contrairement à ce qui est le cas chez la levure. Cependant, des digestions d'ADN génomique provenant de cellules mammifères, par des enzymes de restriction ayant des fréquences de coupure élevées permettent de libérer des TRFs, dont les tailles sont caractéristiques pour les souches de cellules (figure 3). Les télomères de mammifères possèdent aussi une extension terminale 3' simple-brin qui peut être de taille variable (voir plus loin).



Figure 3. Structure d'un télomère humain. Les chromosomes se terminent par ~ 2-50 kpb de répétitions téломériques double-brin, en moyenne. Les digestions de l'ADN génomique avec des enzymes de restriction (HinfI et RsaI) qui coupent fréquemment dans l'ADN non-téломérique permettent de libérer des TRFs caractéristiques des souches de cellules humaines.

Les différentes fonctions associées aux télomères

Le fait de posséder des chromosomes linéaires plutôt que circulaires ne vient pas sans causer de problèmes pour les cellules. En effet, les cellules doivent maintenant faire la différence entre les structures très semblables que sont les cassures d'ADN double-brin (qui doivent être réparées pour la survie de la cellule) et les extrémités des chromosomes (qui doivent à tout prix éviter les fusions pour maintenir la stabilité du génome). De plus, les cellules doivent aussi résoudre le problème relatif à la réplication de leurs extrémités, qui était inexistant avec des chromosomes circulaires. Ces nouveaux problèmes sont en partie réglés par la présence des télomères aux extrémités des chromosomes, qui ont des fonctions importantes à jouer dans la cellule. Une de ces fonctions est de protéger les extrémités des chromosomes contre plusieurs événements possibles dans la cellule (dégradation, recombinaison, ligation). Une autre fonction essentielle des télomères est de différencier les extrémités des chromosomes des cassures d'ADN double-brin. Cette fonction cruciale implique un mode de protection particulier aux extrémités des chromosomes, et elle est souvent intimement liée à la structure de l'ADN terminal. Enfin, les télomères possèdent une autre fonction essentielle qui est de permettre la réplication complète des chromosomes. Les sections suivantes expliquent en détail les fonctions accomplies par les télomères. De plus, les télomères jouent un rôle dans l'organisation nucléaire et sont requis pour la ségrégation des chromosomes lors de la méiose.

Les différentes fonctions associées aux télomères : la protection des chromosomes

La fonction de protection des chromosomes est intimement associée à la structure terminale et à la composition riche en résidus guanines des extrémités des chromosomes.

Étant donné les variations possibles de ces deux éléments entre les espèces, les modes de protection peuvent varier entre les espèces et parfois même, entre les chromosomes d'une même espèce. Par contre, une donnée demeure invariable, c'est-à-dire l'importance de ces éléments pour la protection des extrémités de chromosomes par la formation d'un capuchon protecteur (Blackburn 2000).

Un premier mode de protection des extrémités consiste en la liaison directe des extensions simple-brin par des protéines ou complexes de protéines spécifiques, empêchant ainsi l'accès des exonucléases à l'ADN. Chez le cilié *Oxytricha nova*, les courtes extensions G-riches terminales de l'ADN du macronoyau sont liées par la protéine TEBP (telomere end binding protein). Cette protéine agit sous forme de complexe hétérodimérique α/β . Il a été montré que ce complexe peut lier l'extension G-riche et empêcher la dégradation par la nucléase Bal 31 *in vitro* (Gottschling et Zakian 1986; Price et Cech 1987; Gray et al. 1991). La structure du cristal des protéines d'*O. nova* avec un ADN simple-brin a révélé la présence de 3 repliements de type OB (oligonucleotide/oligosaccharide-binding folds, i.e. OB folds) qui collaborent pour former une structure qui lie l'ADN (Horvath et al. 1998). Chez *Euplotes crassus*, une sous-unité α similaire à celle retrouvée chez *O. nova* lierait aussi l'ADN télomérique, et elle serait également responsable de la protection des courtes extensions G-riches. Pour cet organisme, il n'y aurait pas présence de sous-unité β (Wang et al. 1992). Le mode de protection des extrémités chez le cilié *Tetrahymena* demeure inconnu jusqu'à présent. Aucun complexe protecteur n'a encore été identifié chez cet organisme, même si celui-ci possède également de courtes extensions télomériques comme observé chez *O. nova*.

Chez la levure *S. cerevisiae*, le gène essentiel CDC13 encode une protéine de 924 acides aminés qui a des fonctions de protection analogues à celles du complexe TEBP α/β décrit auparavant (Garvik et al. 1995). La protéine Cdc13p lie les répétitions télomériques TG₁₋₃ simple-brin *in vitro* et lie les télomères *in vivo* (Lin et Zakian 1996; Nugent et al. 1996; Bourns et al. 1998). Le domaine de liaison à l'ADN est compris entre les acides aminés 557 à 694 (Hughes et al. 2000). Il a également été montré que la protéine Cdc13p lie avec une haute affinité un substrat télomérique d'au moins 11 nucléotides, considéré comme un site consensus minimal pour la liaison (Hughes et al. 2000).

Le rôle dans la protection des extrémités de la protéine Cdc13p a été montré par l'utilisation d'un allèle thermosensible (*cdc13-1*) (Garvik et al. 1995). Chez la levure, des mutations qui rendent des protéines essentielles ayant un rôle dans la progression du cycle cellulaire thermosensibles peuvent aussi mener à des arrêts de croissance lors de phases particulières du cycle cellulaire, qui sont dépendants des protéines de checkpoint. Les cellules possédant l'allèle *cdc13-1* arrêtent en phase G2 du cycle cellulaire lorsqu'elles sont poussées à la température restrictive, et cet arrêt est dépendant de la protéine de checkpoint Rad9p (Garvik et al. 1995). Lors de l'analyse de l'ADN de ces cellules, il a été montré qu'il y avait accumulation de grandes quantités d'ADN simple-brin aux extrémités des chromosomes, et ce simple-brin correspondait en fait au brin télomérique G-riche (Garvik et al. 1995). Dans les souches *cdc13-1*, il y a donc dégradation extensive du brin télomérique C-riche, sur plusieurs kilobases (Garvik et al. 1995), ce qui suggère que la protéine Cdc13p est requise pour la protection des extrémités contre la dégradation (Garvik et al. 1995). Un modèle pouvant expliquer le rôle de Cdc13p a été mis de l'avant par le groupe de Vicki Lundblad, et suggère que

Cdc13p se lie au simple-brin G-riche télomérique, empêchant ainsi la dégradation du brin complémentaire par les exonucléases (Nugent et al. 1996).

L'interaction de la protéine Cdc13p avec les protéines Stn1p et Ten1p suggère que la protection des extrémités des chromosomes soit dépendante des complexes formés par ces trois protéines (Grandin et al. 1997; Grandin et al. 2001; Pennock et al. 2001). Il a été montré par double-hybride que les protéines Cdc13p et Stn1p interagissent entre elles (Grandin et al. 1997). La perte de Stn1p dans les cellules mène à un arrêt cellulaire en phase G2/M qui est dépendant de la protéine Rad9p (Grandin et al. 1997) et à une accumulation d'ADN simple-brin dans les régions télomériques (Grandin et al. 2001), les mêmes phénotypes que ceux observés avec l'allèle *cdc13-1*. La protéine Ten1p s'associe également avec les protéines Cdc13p et Stn1p (Grandin et al. 2001); elle lie la protéine Stn1p *in vivo* et s'associe avec la protéine Cdc13p dans des essais de double-hybride (Grandin et al. 2001). Il est également observé que des mutations dans le gène TEN1 mènent à des arrêts cellulaires en phase G2/M, dépendants de la protéine Rad9p (Grandin et al. 2001). Ces mutants de TEN1 accumulent aussi de l'ADN simple-brin dans les régions télomériques (Grandin et al. 2001). De plus, il a été montré que la création d'une protéine chimérique entre le domaine de liaison à l'ADN de Cdc13p (DBD_{CDC13}) et Stn1p permet la survie d'une souche de levure *cdc13Δ* (Pennock et al. 2001). Ces évidences suggèrent que Stn1p joue un rôle primordial dans la protection des extrémités des chromosomes, et que Cdc13p permet le positionnement adéquat de cette protéine aux extrémités télomériques (Pennock et al. 2001). En plus de son rôle dans la protection des télomères, Cdc13p a également un rôle important dans la réplication des télomères. Les différentes analyses de séquences effectuées entre les protéines liant le simple-brin de

différentes espèces et la protéine Cdc13p de levure n'ont pas permis d'établir de liens fonctionnels entre les protéines. Cependant, des analyses de la structure des protéines ont permis de placer Cdc13p dans la famille des protéines impliquées dans la protection des extrémités. En effet, l'analyse de la structure du domaine de liaison à l'ADN de Cdc13p avec l'ADN télomérique révèle la présence d'un domaine de type OB (OB fold) (Mitton-Fry et al. 2002; Mitton-Fry et al. 2004). Ce domaine de type OB possède une structure similaire à ceux retrouvés chez les ciliés, malgré le fait qu'il n'y ait pas de similarité au niveau de la séquence (Mitton-Fry et al. 2002; Mitton-Fry et al. 2004). Donc, cette homologie de structure entre les protéines TEBP d'*O. nova* et Cdc13p de *S. cerevisiae* porte à croire que, malgré la divergence de leurs séquences, ces protéines pourraient partager un ancêtre commun (Mitton-Fry et al. 2002). Il est maintenant connu que les protéines impliquées dans la protection des extrémités des chromosomes chez différentes espèces possèdent des domaines de types OB, ce qui démontre la conservation du mécanisme de protection des extrémités à travers le règne eucaryote.

Chez la levure *S. pombe*, une comparaison de séquences à partir de banques de données a permis d'identifier la protéine Pot1p (protection of telomeres), qui contient des régions d'homologie avec la sous-unité α du complexe TEBP d'*O. nova* (Baumann et Cech 2001). La protéine Pot1p lie les séquences télomériques G-riches (Baumann et al. 2002). La délétion de Pot1p a des effets dramatiques sur la stabilité des télomères et des chromosomes; après seulement 10 générations suite à la délétion de la protéine les cellules cessent de se diviser, il y a une perte immense d'ADN télomérique (dégradation d'environ 5 kb) et la seule façon dont les cellules peuvent survivre est la circularisation de leurs chromosomes (Baumann et al. 2002). Il a été montré que la protéine Pot1p de *S.*

pombe peut lier un substrat d'ADN télomérique aussi court que 6 nucléotides (Lei et al. 2002), et que le domaine de liaison à l'ADN de la protéine est de type OB fold (Lei et al. 2003). Grâce aux études génétiques effectuées chez *S. pombe*, plusieurs évidences montrent que le groupe de protéines ayant une homologie de séquences (la sous-unité α du complexe TEBP d'*O. nova*, la protéine α de *Euplotes* et la protéine Pot1p de *S. pombe*) joue un rôle crucial dans la protection des extrémités chromosomiques et la prévention de la dégradation rapide des télomères *in vivo* (Baumann et al. 2002). L'analyse de séquences a permis d'identifier un homologue humain de la protéine Pot1p (appelé hPot1p) (Baumann et al. 2002), qui est décrit dans la prochaine section. Une protéine Pot1p a aussi été identifiée chez les cellules de poulet, qui est capable de lier un substrat d'ADN télomérique minimal de 12 nucléotides (Wei et Price 2004) et une protéine homologue a aussi été identifiée chez les plantes (Baumann et Cech 2001), (Baumann et al. 2002). Au moins un domaine de type OB (OB fold) présent dans le domaine de liaison à l'ADN de la protéine Pot1p du poulet est requis pour la liaison à l'ADN télomérique (Wei et Price 2004).

Chez les cellules humaines, la protéine hPot1p a été identifiée par homologie de séquences avec la protéine Pot1p de *S. pombe* et la sous-unité α du complexe TEBP d'*O. nova* (Baumann et al. 2002). Cette protéine se lie de façon spécifique à l'ADN télomérique humain du brin G-riche, et ne lie pas le brin C-riche ni l'ADN télomérique double-brin (Baumann et al. 2002). Il a été montré que le substrat télomérique minimal pour la liaison de hPot1p est 9 nucléotides (Loayza et al. 2004). La délétion du domaine de type OB (OB fold) de hPot1p abolit la liaison de la protéine à l'ADN télomérique (Loayza et al. 2004). La protéine hPot1p a aussi une fonction au niveau du recrutement

de la télomérase, comme son homologue fonctionnel Cdc13p chez la levure *S. cerevisiae* (Colgin et al. 2003; Armbruster et al. 2004). De plus, il est maintenant clair que hPot1p interagit avec de nombreuses autres protéines impliquées dans la protection des télomères chez l'humain, menant à la formation d'un complexe protéique (appelé télosome) aux extrémités télomériques des cellules humaines (Liu et al. 2004a; Liu et al. 2004b; Ye et al. 2004a; Ye et al. 2004b).

Un second mode de protection des extrémités des chromosomes a été mis en évidence chez les cellules de mammifères en 1999. Par microscopie électronique, il a été

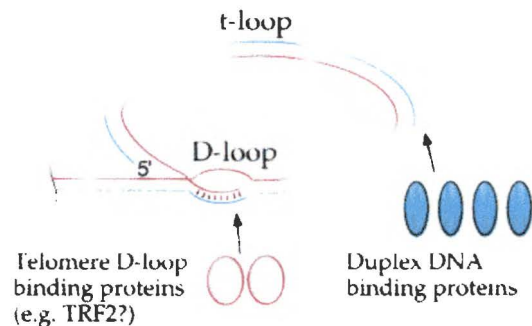


Figure 4. Structure des t-loop et D-loop. L'ADN télomérique se replie sur lui-même, formant ainsi la structure de boucle nommée t-loop (telomere loop). L'extension 3' simple-brin est introduite dans la portion double-brin d'ADN et forme la D-loop (displacement loop), ce qui stabilise la structure. Tiré de Greider (1999).

montré que les chromosomes de mammifères peuvent terminer en de larges boucles terminales, qui ont été nommées t-loops (Griffith et al. 1999). La formation de ces boucles nécessite la présence d'une extension télomérique simple-brin assez longue (Griffith et al. 1999), et est également dépendante de la protéine télomérique TRF2, retrouvée liée à la base de la t-loop et qui facilite la formation de

ces structures dans des essais *in vitro* (Griffith et al. 1999; Yoshimura et al. 2004) (figure 4). Le mécanisme proposé pour expliquer la formation de ces structures terminales est l'invasion de l'extrémité simple-brin terminale dans la portion télomérique double-brin (Griffith et al. 1999). L'extrémité télomérique G-riche se retrouve ainsi appariée avec une séquence télomérique C-riche située à l'intérieur du télomère (Griffith et al. 1999). Ce mécanisme déplace une fraction de la portion télomérique double-brin G-riche qui

devient simple-brin et qui forme ce qui est appelé la D-loop (displacement loop) (Griffith et al. 1999). Il a été proposé que cette structure serve à cacher l'extrémité terminale du chromosome, et à la différencier d'une cassure d'ADN double-brin (Griffith et al. 1999). En accord avec ce modèle, il a été montré que la surexpression d'une version dominante négative de la protéine TRF2 mène à une perte des extensions G-riches simple-brin et que des fusions entre les télomères sont induites (van Steensel et al. 1998). De plus, l'absence d'une protéine TRF2 normale mène à une apoptose qui est dépendante des protéines de checkpoint p53 et ATM, ce qui confirme que des dommages importants sont détectés par la cellule si l'on inhibe la protéine TRF2 (Karlseder et al. 1999).

Ce nouveau mode de protection sous forme de t-loops a aussi été mis en évidence chez *Oxytricha fallax* et *Trypanosoma brucei* (Murti and Prescott 1999; Munoz-Jordan et al. 2001), ce qui suggère que ces structures peuvent être conservées entre les espèces. Une caractéristique des télomères des ciliés *O. fallax* et *T. brucei* est la présence de longues extensions simple-brin, ce qui semble être un pré-requis pour la formation des t-loops (Munoz-Jordan et al. 2001; Murti and Prescott 1999; Griffith et al. 1999). Les différentes protéines requises et impliquées dans la formation de ces structures n'ont pas encore été identifiées pour les ciliés. Des structures de t-loops n'ont pas encore été mises en évidence chez *S. cerevisiae*, car les extensions terminales présentes aux télomères sont plutôt courtes (Larrivée et al. 2004), sauf pour une courte période du cycle cellulaire où elles deviennent plus longues suite à leur réplication (Wellinger et al. 1993b). Il est cependant impossible de dire avec certitude que ces structures sont inexistantes. Certaines protéines de la levure pourraient permettre la formation ces structures malgré la présence de courtes extensions télomériques. Chez *S. cerevisiae*, une structure alternative, mais

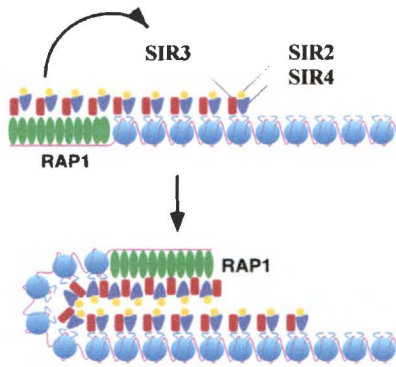


Figure 5. Modèle pour la formation de l'hétérochromatine télomérique chez *S. cerevisiae*. Rap1p se lie aux répétitions télomériques. Les protéines SIR interagissent avec Rap1p et l'histone H4. Elles peuvent aussi interagir entre elles et amorcer le repliement de la structure. Les cercles bleus illustrent les nucléosomes. Tiré de Grunstein (1997).

similaire, de repliement des télomères a été proposée par le groupe de Grunstein. Cette structure implique la participation et l'interaction des protéines de l'hétérochromatine télomérique, et pourrait s'avérer être la façon dont *S. cerevisiae* protège ses extrémités de chromosomes (Grunstein 1997; de Bruin et al. 2000; de Bruin et al. 2001) (figure 5). Récemment, des évidences *in vitro* suggèrent que la protéine Taz1p de *S. pombe* puisse aider à la formation de ces structures sur des substrats télomériques ayant une extension simple-brin,

suggérant que ces structures puissent être présentes chez la levure de fission (Tomaska et al. 2004b). Ces évidences renforcent l'hypothèse que ce mode de protection est conservé entre les eucaryotes au cours de l'évolution.

Les différentes fonctions associées aux télomères : la réplication complète des chromosomes

En plus de devoir trouver une façon d'assurer la protection de leurs extrémités, les eucaryotes vivant avec des chromosomes linéaires doivent faire face à un deuxième problème, concernant la réplication de l'extrémité des chromosomes. En effet, les ADN

polymérase conventionnelles sont incapables de répliquer la toute fin des molécules linéaires d'ADN (Watson 1972; Olovnikov 1973). Les ADN polymérase nécessitent une amorce ARN pour initier la réplication de l'ADN et synthétisent l'ADN uniquement dans la direction 5' vers 3'. Ainsi, lors de la réplication des extrémités terminales des chromosomes par la machinerie conventionnelle de réplication, l'enlèvement de la dernière amorce ARN placée sur l'ADN télomérique laisse un trou à l'extrémité 5' du

brin C-riche nouvellement synthétisé (figure 6).

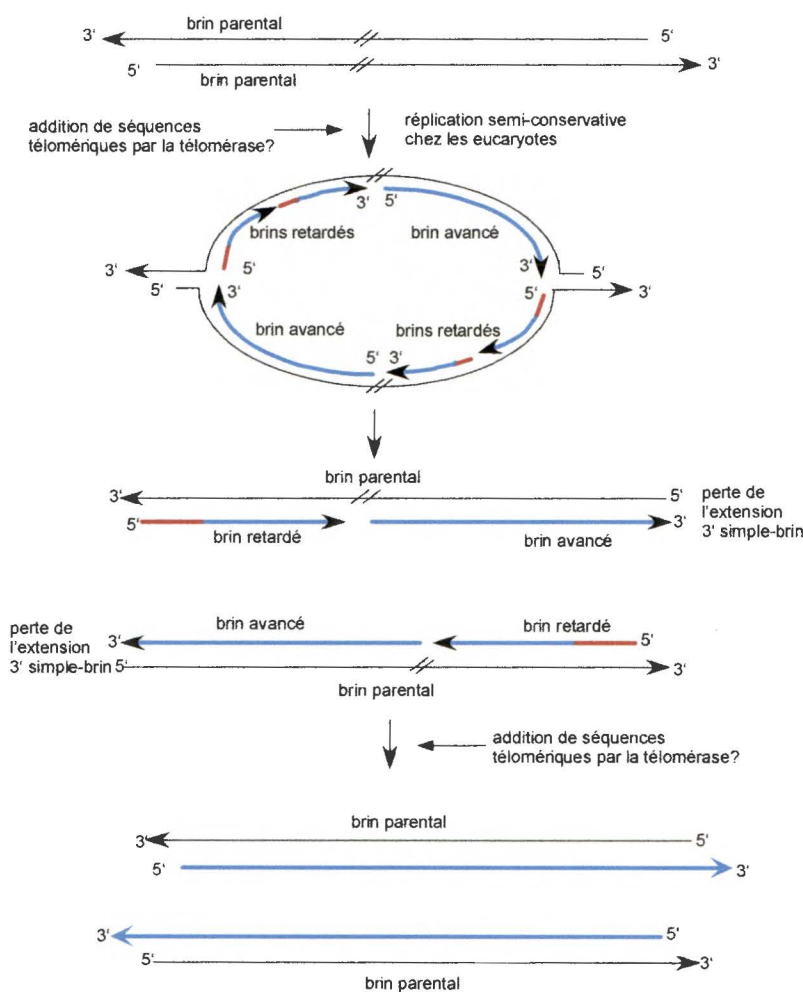


Figure 6. Schéma de la réplication de l'ADN et du problème du brin avancé lors de la réplication des télomères (end-replication problem). L'ADN parental (noir) est répliqué (ADN nouvellement synthétisé en bleu) par un mécanisme semi-conservatif. Les flèches indiquent la polarité de l'ADN. Suite à la réplication, il y a perte de l'extension 3' simple-brin au brin avancé. La création subséquente du substrat pour la télomérase ne peut être possible qu'avec l'action d'une exonucléase. Cependant, à l'extrémité du brin retardé, l'enlèvement de la dernière amorce ARN (rouge) permet la création d'une courte extension 3' simple-brin et la perte de séquences terminales. Tiré de LeBel et Wellinger (2004).

Si aucun mécanisme n'est enclenché par la cellule pour contrer cette perte des séquences terminales, il y aura perte des séquences télomériques au fil des cycles de réplication. Chez les humains, on observe effectivement un raccourcissement des

télomères dans la plupart des cellules somatiques au fil des divisions cellulaires. Comme ce phénomène de perte de séquences télomériques se produit à chacune des rondes de réplication de l'ADN, il mène en bout de ligne à une trop grande perte de séquences télomériques et à la mort cellulaire. Puisque les télomères sont essentiels pour assurer la protection des chromosomes, la cellule doit trouver un moyen d'assurer la réplication complète des télomères sans subir la perte des séquences terminales.

Pour la plupart des eucaryotes, le problème de réplication complète des extrémités des chromosomes est résolu par l'activation d'une enzyme spécifique appelée télomérase. Cette enzyme a d'abord été identifiée chez le cilié *Tetrahymena* en 1985 (Greider et Blackburn 1985; Greider and Blackburn 1987). Il a été montré qu'elle peut ajouter des

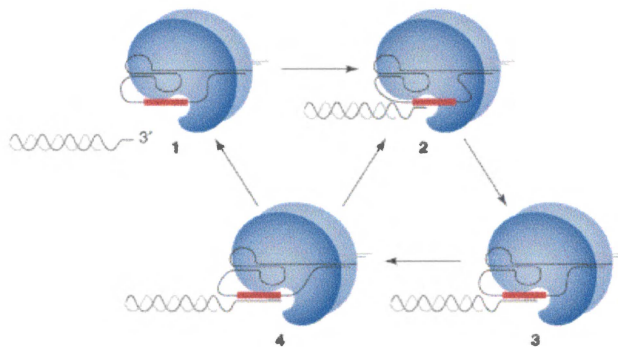


Figure 7. Modèle de l'élongation des télomères par la télomérase. Différentes étapes de la réaction d'élongation sont décrites ici. La télomérase est représentée par la sphère bleue et la sous-unité ARN est représentée par la structure en noir. La région de l'ARN utilisée comme matrice est en rouge. Le stade 1 montre l'enzyme et le substrat télomérique comme deux entités séparées. Le stade 2 illustre la liaison, par complémentarité de bases, entre la région matrice de l'ARN de la télomérase et la portion simple-brin des télomères. Le stade 3 représente un cycle d'élongation des télomères par la télomérase et l'ajout de séquences télomérique. Lorsque l'élongation des télomères a lieu jusqu'au bout de la région matrice de la sous-unité d'ARN (stade 4), il peut y avoir dissociation entre l'enzyme et le substrat allongé (retour au stade 1) ou translocation de l'enzyme (retour au stade 2) et élongation successive. Modèle tiré de Kelleher et al. (2002).

répétitions télomériques aux extrémités des chromosomes naturels, et ainsi empêcher leur raccourcissement graduel (revue dans (Collins 1996; Collins 1999; Harrington 2003)). Cette enzyme est une ribonucléoprotéine composée, chez la majorité des organismes, d'une sous-unité catalytique, d'une matrice ARN et de protéines associées. Son mode d'action est celui d'une transcriptase inverse, et elle peut allonger le brin télomérique G-riche (substrat simple-brin) en

utilisant une portion de l'ARN associé à l'enzyme pour dicter l'ajout de séquences (Greider et Blackburn 1985; Greider et Blackburn 1987; Lingner et Cech 1996) (figure 7).

D'abord découverte et identifiée chez *Tetrahymena*, la télomérase est maintenant identifiée chez plusieurs organismes, dont certains autres ciliés (Zahler et al. 1991; Shippen-Lentz et Blackburn 1989), l'humain (Morin 1989), la souris (Prowse et al. 1993), la levure (Cohn et Blackburn 1995) et la plante (Fitzgerald et al. 1996). Suite à différentes approches menant à la purification de l'holoenzyme télomérase, de plus en plus de protéines sont maintenant connues comme faisant partie ou étant associées à la télomérase, chez plusieurs organismes (revue dans (Collins et Mitchell 2002; Cech 2004; Smogorzewska and de Lange 2004)).

Chez la levure *S. cerevisiae*, certains criblages génétiques ont permis l'identification de 5 gènes essentiels au maintien des séquences télomériques par la télomérase. Le gène *EST2* (ever shorter telomere) encode la sous-unité catalytique de l'enzyme alors que les gènes *EST1*, *EST3* et *CDC13/EST4* encodent des facteurs de recrutement et de régulation de l'enzyme (Lundblad and Szostak 1989), (Lendvay et al. 1996), (Lin et Zakian 1996; Nugent et al. 1996; Virta-Pearlman et al. 1996; Counter et al. 1997; Lingner et al. 1997a; Lingner et al. 1997b). Le gène *TLC1* (telomerase component 1) encode la composante ARN de la télomérase (Singer et Gottschling 1994). Il a été observé que seulement Est2p et TLC1 sont essentiels pour reconstituer une activité télomérasique *in vitro* (Lingner et al. 1997b).

Des souches de levure qui possèdent une délétion d'un seul de ces 5 gènes essentiels sont incapables de compléter la réplication de leurs télomères. Alors, les

séquences télomériques raccourcissent d'environ 3 à 5 pb par génération, ce qui conduit ultimement à une perte de viabilité après environ 50 à 70 générations et à la sénescence (Lundblad et Szostak 1989; Singer et Gottschling 1994). Cependant, il existe des mécanismes indépendants de la télomérase qui peuvent assurer le maintien des télomères chez la levure. Dans une population de levures en sénescence, certaines cellules vont arriver à contourner ce problème et devenir des survivants (Lundblad et Blackburn 1993). Dans ces survivants, les répétitions télomériques sont maintenues par un processus de recombinaison dépendant de la protéine Rad52p (Lundblad et Blackburn 1993; Le et al. 1999; Teng et Zakian 1999; Teng et al. 2000). Les survivants sont classifiés en deux groupes distincts, possédant chacun des caractéristiques propres (Lundblad et Blackburn 1993). Les survivants de type I possèdent des télomères très courts et montrent une amplification des éléments Y' (Lundblad et Blackburn 1993). Ce type de survivant dépend de la protéine Rad51p, qui permet l'échange de brins (Le et al. 1999), et des protéines Rad55p, Rad54p et Rad57p, en plus de la protéine Rad52p. Les survivants de type II possèdent des télomères très longs et hétérogènes pouvant atteindre jusqu'à 12 kb (Lundblad et Blackburn 1993; Le et al. 1999; Teng et al. 2000). La génération de ce type de survivants dépend de la protéine Rad50p (Le et al. 1999) et des protéines Xrs2p et Mre11p, ainsi que de la protéine Rad52p.

Chez les mammifères, le maintien des télomères s'effectue aussi par la télomérase. La sous-unité catalytique et la composante ARN de l'enzyme humaine ont été clonées (Feng et al. 1995; Harrington et al. 1997; Meyerson et al. 1997; Nakamura et al. 1997), et de plus en plus de nouvelles protéines sont découvertes comme faisant partie de l'holoenzyme (revue dans (Smogorzewska et de Lange 2004)). Cependant, chez les

mammifères, l'activation de l'enzyme est différente de ce qui est connu chez la levure. En effet, dans les cellules somatiques humaines normales, la télomérase s'avère présente à de bas niveaux et est très peu active (Masutomi et al. 2003), mais elle est activée dans les cellules germinales, dans la plupart des lignées de cellules transformées et dans 85% des cellules cancéreuses primaires (Kim et al. 1994). Certaines lignées cellulaires humaines immortalisées peuvent maintenir leurs télomères en absence de télomérase, par un mécanisme alternatif (alternative lengthening of telomeres; ALT) qui serait similaire à celui utilisé pour la formation des survivants chez la levure (Bryan et al. 1995), (Bryan et al. 1997a), (revue dans (Reddel 2003)).

L'implication des télomères dans la détermination du futur des cellules

Étant donné le lien existant entre la longueur des télomères, l'activation de la télomérase et l'immortalisation des cellules, des modèles ont vu le jour pour tenter d'expliquer et de mieux comprendre ces différents phénomènes cellulaires. À partir de 1990, un des modèles les plus populaires à ce sujet était celui « d'horloge mitotique » ou de la limite de Hayflick (Harley et al. 1990; Harley 1991). Ce modèle était supporté par plusieurs observations, la principale étant que dans des cellules somatiques normales, la télomérase est inactive et les télomères raccourcissent au fil des divisions cellulaires (Harley et al. 1990). Étant donné que les télomères sont essentiels à la stabilité du génome, lorsque ceux-ci sont trop érodés après de nombreuses rondes de divisions, la cellule entre en sénescence pour ne pas propager de défauts possibles lors de divisions subséquentes. Ce serait le raccourcissement progressif de la longueur des télomères avec le nombre de divisions dans les cellules somatiques qui limite la capacité de prolifération de la plupart des cellules humaines. Ce modèle stipule donc que la longueur des

télomères *per se* agirait comme élément critique dans le destin d'une cellule (Harley 1991; Harley et al. 1990). En accord avec ce modèle est l'observation de la relation du potentiel réplicatif des cellules et de la longueur des télomères (Allsopp et al. 1992), qui dit que la longueur des télomères, et donc le potentiel de réplication, diminue avec l'augmentation des individus en âge, chez les humains et les souris (Allsopp et al. 1992; Prowse et Greider 1995). Cependant, certaines cellules peuvent échapper à la sénescence et à la mort cellulaire et environ 85% d'entre elles le font en activant la télomérase (Kim et al. 1994) et environ 15% des cellules trouvent un moyen alternatif pour maintenir leurs télomères et échapper à la mort cellulaire (Bryan et al. 1995; Bryan et al. 1997a; Reddel et al. 1997). Il a été montré expérimentalement que l'inactivation de la télomérase cause le raccourcissement des télomères et la sénescence cellulaire, alors que l'activation de la télomérase permet le rallongement des télomères et l'immortalisation des cellules (Counter et al. 1992; Bodnar et al. 1998). La réduction générale de la taille des télomères peut être vue comme un mécanisme de suppression des tumeurs, permettant la mort des cellules avant le développement d'aberrations. Plus spécifiquement, il a été montré par le groupe de Greider que l'activation des réponses cellulaires peut être dépendante d'un télomère se retrouvant plus court qu'une certaine taille critique (Hemann et al. 2001). Il a été montré que la perte de fonction associée aux télomères arrive préférentiellement aux télomères les plus courts de la cellule, et que ce sont également ce télomères les plus courts qui sont allongés de façon préférentielle par l'enzyme télomérase (Hemann et al. 2001). L'activation de la télomérase et le maintien de la longueur des télomères semblent un pré-requis pour la croissance continue et la progression de tumeurs. Cependant, certaines observations ne peuvent être incorporées à ce modèle, et soulèvent le doute

quant à son universalité. Entre autres, il y a des cas où les cellules entrent en sénescence sans réduction détectable de la longueur de leurs télomères, et d'autres cas où les cellules qui sont en sénescence ont des télomères plus longs que des cellules qui sont encore en division active (Cristofalo et al. 1998; MacKenzie et al. 2000; Frenck et al. 1998; Weng et al. 1998; revue dans (Cristofalo et al. 2004). Dans ces différents cas, il est possible que les mécanismes soient induits par un ou des télomère(s) court(s) présent(s) dans la cellule, dont la taille critique permettant l'induction du signal est masquée lors des analyses par la taille des autres télomères (Hemann et al. 2001).

Pour rallier ces observations, un nouveau modèle a été mis de l'avant en 2000. Ce modèle propose plutôt que le destin des cellules est décidé par l'état de protection des télomères (Blackburn 2000). Ce serait l'intégrité du capuchon protecteur formé par l'ADN télomérique et les protéines liées aux télomères qui serait important pour la continuation des divisions cellulaires (Blackburn 2000). Des changements dans certains aspects cruciaux relatifs à la protection des télomères activeraient la sénescence répllicative, et non seulement la diminution de la longueur des télomères (Blackburn 2000). Dans ce modèle (figure 8), les extrémités des chromosomes sont protégées par le capuchon télomérique. Le raccourcissement progressif des télomères suite aux divisions cellulaires ou l'absence d'activité télomérase peut mener à la perte de ce capuchon. Cependant, l'activation de la télomérase ou du processus de recombinaison homologue peut ramener les télomères à un état normal qui permet leur protection, et la poursuite des divisions cellulaires. Si aucun moyen n'est pris par la cellule pour protéger les extrémités dégarnies de leur capuchon télomérique, des signaux spécifiques sont envoyés à la cellule et celle-ci entre en sénescence.

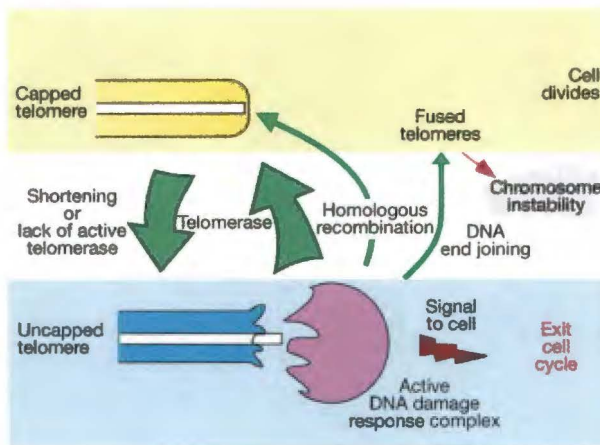


Figure 8. Changements dans la protection des télomères. Si un télomère perd son capuchon protecteur, la réponse appropriée est l'activation de la télomérase (majoritaire) ou la recombinaison homologue. Ceci permet de protéger et/ou rallonger le télomère et continuer les divisions cellulaires. Il peut aussi y avoir des fusions entre télomères ce qui les rend instables. Si aucun mécanisme n'est enclenché suite à la déprotection des télomères, la réponse d'une cellule normale est d'arrêter la division cellulaire ou, pour certaines cellules mammifères, d'enclencher l'apoptose. Tiré de Blackburn (2000).

Plusieurs facteurs sont à considérer pour assurer la protection complète des télomères dans les cellules, dont, l'activité de l'enzyme télomérase, la longueur des télomères, les complexes formés entre les protéines spécifiques et l'ADN télomérique double-brin, les complexes protéiques présents à l'extrémité 3' simple-brin de l'ADN télomérique et l'intégrité de la portion simple-brin présente à l'extrémité des télomères (Blackburn 2000). En effet, l'extension 3' simple-brin a un rôle majeur dans la formation du capuchon protecteur et dans la protection même des télomères. Dans la section suivante, l'extension 3' simple-brin de différentes espèces est décrite, ainsi que les évidences sur son rôle dans la protection de l'intégrité des télomères.

La structure 3' terminale des extrémités des chromosomes et son importance

La structure terminale exacte des chromosomes de différentes espèces est étudiée en détails depuis plusieurs années (tableau 2). Cette structure comportant l'extension 3' simple-brin s'avère cruciale pour permettre les différentes fonctions qui sont associées aux télomères. Elle est détectée chez tous les organismes étudiés, prouvant ainsi qu'elle est une caractéristique bien conservée entre les espèces et qu'elle peut donc avoir une

importance fonctionnelle. Les premières expériences concernant la détermination de la séquence et de la taille de l'extension simple-brin télomérique ont été faites chez les ciliés *Tetrahymena*, *Oxytricha* et *Euplotes*. La taille de l'extension simple-brin chez ces organismes a été estimée entre 12 et 16 nucléotides (Klobutcher et al. 1981).

Le groupe de Carolyn Price s'est intéressé plus en détails à la dynamique des télomères du cilié *Tetrahymena thermophila*. Avec une technique élégante d'extension d'une amorce qui a été liguée au substrat (ligation-mediated primer extension), il a été montré que la taille de l'extension 3' télomérique est très strictement contrôlée chez *Tetrahymena*, et plus de 80% des extensions ont une taille de 14-15 nucléotides ou de 20-21 nucléotides (Jacob et al. 2001). De plus, la plupart de ces extensions possède une séquence précise, soit 5'-TTGGGGT-3', et les extensions sont détectables durant tout le cycle cellulaire (Jacob et al. 2001). Il est également montré que la taille des extensions télomériques peut varier lorsque les cellules sont soumises à certaines restrictions nutritionnelles ou lorsqu'elles sortent du cycle cellulaire (Jacob et al. 2001). Certaines évidences suggèrent également que les deux extrémités d'un chromosome possèdent une extension télomérique simple-brin (Jacob et al. 2001). Ceci implique que les deux télomères présents sur un même chromosome du macronoyau de *Tetrahymena* peuvent être empaquetés dans des structures protectrices semblables. La présence de courtes extensions porte à croire que la protection des extrémités des chromosomes s'effectue par la liaison de protéines spécifiques qui peuvent reconnaître l'extension simple-brin et former des complexes protecteurs. La présence de t-loops semble improbable chez *Tetrahymena*, étant donné le besoin de longues extensions simple-brin pour en permettre la formation.

Chez la levure *S. cerevisiae*, la taille des extensions télomériques simple-brin semble dépendre des phases du cycle cellulaire. En effet, les premiers résultats à ce sujet montrent que les télomères d'un plasmide linéaire présent dans la levure acquièrent des extensions simple-brin transitoires d'au moins 30 bases, à la fin de la phase S (Wellinger et al. 1993a; Wellinger et al. 1993b). Cette période du cycle cellulaire corrèle avec la période de réplication de l'ADN télomérique chez la levure, tard en phase S (Wellinger et al. 1993a; Marcand et al. 2000). Les résultats montrent également que les extensions simple-brin sont présentes aux deux extrémités d'un plasmide linéaire, par la formation de molécules circulaires (CFP, circular form of the plasmid) dépendantes de ces extensions (Wellinger et al. 1993b). Contrairement à ce qui est observé chez les ciliés, la séquence 3' terminale de la levure peut être différente pour chaque télomère. Cette séquence correspond toujours à la séquence consensus TG₁₋₃, mais le nombre de résidus guanines présents dans chacune des répétitions peut varier (Forstemann et Lingner 2001). Des limites de la technique d'hybridation dans le gel (in-gel hybridization) (Dionne et Wellinger 1996; McElligott et Wellinger 1997) ne permettent pas la détermination de la structure terminale des extensions télomériques en dehors de la phase S.

Des améliorations de cette méthode et le développement d'une méthode alternative pour la mesure du simple-brin télomérique chez la levure ont permis de montrer la structure des télomères de levure en phase G1 du cycle cellulaire. Grâce à l'arrêt de cellules possédant un plasmide linéaire de haute copie en phase G1 du cycle cellulaire, il est possible de détecter des extensions télomériques simple-brin, qui se retrouvent concentrées en une bande précise et visible dans le gel (Larrivée et al. 2004). D'autres améliorations de la méthode d'hybridation dans le gel ont permis de déterminer

que la taille des extensions télomériques chez la levure en phase G1 est plus courte que 16 bases (Larrivée et al. 2004). Cette détermination précise ne pouvait pas être faite auparavant avec les techniques disponibles.

Une méthode d'extension d'amorce modifiée sur des extensions télomériques réelles a permis de conclure que la taille des extensions simple-brin chez des levures de type sauvage en G1 est de 12 à 14 bases (Larrivée et al. 2004). De plus, il est montré que la vaste majorité des télomères de levure possèdent ces extensions télomériques simple-brin en G1 (Larrivée et al. 2004). La présence, la taille et la dynamique de génération de ces extensions sont déterminées par certaines protéines importantes ayant des implications multiples aux télomères. Il a été montré que la présence des extensions télomériques est dépendante de la réplication de l'ADN et du passage d'une fourche de réplication (Dionne et Wellinger 1998), et que ces extensions peuvent être détectées dans des cellules qui n'expriment pas la télomérase (Dionne et Wellinger 1996). Cette évidence suggère qu'il y aurait une étape de dégradation du brin C-riche pour permettre la création de l'extension G-riche en absence de télomérase (Dionne et Wellinger 1996), (Wellinger et al. 1996). Étant donné les différences observées entre les ciliés et la levure, il semble que la dynamique des extensions simple-brin au cours du cycle cellulaire est variable selon les espèces.

L'analyse des télomères de plantes a permis de mettre en évidence deux structures distinctes de leurs extensions télomériques simple-brin (Riha et al. 2000). Grâce à l'utilisation d'une méthode appelée PENT (primer extension/ nick translation) d'abord développée pour l'analyse des télomères de cellules humaines (Makarov et al. 1997) (voir plus loin), la longueur des extensions simple-brin a été estimée entre 20 et 30 nucléotides,

pour environ 50% de la population des télomères (Riha et al. 2000). Les autres télomères posséderaient des extensions d'environ 12 nucléotides (Riha et al. 2000). Les observations suggèrent que les télomères de plantes peuvent avoir deux terminaisons possibles à la fin d'un chromosome, ce qui donne une configuration asymétrique (Riha et al. 2000). Comme chez la levure, l'extension télomérique G-riche est aussi présente sur les chromosomes de cellules ne possédant pas de télomérase (Riha et al. 2000), ce qui renforce son importance fonctionnelle.

En ce qui concerne les cellules de mammifères et plus particulièrement les cellules humaines, il a été montré que les séquences télomériques et les extensions simple-brin sont beaucoup plus longues que celles décrites auparavant. En effet, les télomères des cellules humaines possèdent des extensions G-riche de plus de 150 nucléotides, qui seraient présentes à toutes les phases du cycle cellulaire (Wright et al. 1997; Makarov et al. 1997; McElligott and Wellinger 1997). Cependant, une controverse survient quant à la distribution de ces longues extensions aux deux extrémités d'un chromosome. En utilisant un essai de PENT (primer extension/nick translation), Makarov et collaborateurs ont observé la présence de longues extensions G-riches sur plus de 80% des extrémités des chromosomes humains (Makarov et al. 1997). De leur côté, le groupe de Wright et collaborateurs a utilisé une technique de purification des chromosomes humains par affinité, et a montré que les chromosomes humains sont asymétriques comme ce fut observé chez les plantes, avec une longue extension d'un côté et une courte extension sur l'autre télomère (Wright et al. 1997). Par microscopie électronique, ils ont évalué la taille des longues extensions à environ 200 ± 75 nucléotides (Wright et al. 1997). Ce résultat semble être en accord avec le modèle de réplication conventionnelle

(figure 6), qui prédit que seulement 50% des extrémités, c'est-à-dire celles produites par la réplication du brin retardé, devraient posséder une extension 3' simple-brin, suite à l'enlèvement de la dernière amorce ARN du dernier fragment d'Okazaki. Les résultats de Makarov et collaborateurs montrant de longues extensions simple-brin détectées aussi à l'extrémité du brin avancé peuvent être expliqués par l'action possible de la télomérase ou d'exonucléases. De telles activités ont déjà été mises en évidence chez la levure *S. cerevisiae* (Wellinger et al. 1996), et pourraient également être actives chez la plante. Il est cependant évident que les extensions simple-brin sont présentes à toutes les extrémités des chromosomes, et qu'elles sont essentielles pour leur protection.

Le groupe de d'Ambrosio et collaborateurs a mis au point une méthode de mesure de la longueur des extensions simple-brin présentes aux télomères de cellules humaines. Cette méthode appelée T-OLA (Telomeric Oligonucleotide Ligation Assay) est basée sur l'hybridation et la ligation d'oligonucléotides complémentaires aux extensions télomériques humaines (Cimino-Reale et al. 2001). Les analyses faites sur différents types de cellules humaines, des fibroblastes, des cellules immortalisées (HeLa) et des lymphocytes (Cimino-Reale et al. 2001) ont permis de mettre en évidence des extensions télomériques variant de 24 nucléotides, la taille minimale détectable par cette nouvelle méthode, jusqu'à 360 nucléotides (Cimino-Reale et al. 2001). Cependant, il est possible que les extensions télomériques soient plus courtes que cette limite de détection. Aussi, avec l'ADN isolé de fibroblastes et de lymphocytes, les extensions télomériques simple-brin peuvent aller jusqu'à plus de 400 nucléotides (Cimino-Reale et al. 2001). Suite à des analyses de quantification, il apparaît que chez tous les types de cellules analysées, 37% des extensions télomériques sont d'une taille située entre 108 et 270 nucléotides (Cimino-

Reale et al. 2001). De plus, pour ce qui est des fibroblastes et des lymphocytes, 56% des extensions G-riches sont de taille inférieure à 90 nucléotides (Cimino-Reale et al. 2001). Cette nouvelle approche montre que des extensions simple-brin beaucoup plus courtes que celles détectées auparavant sont présentes aux extrémités des chromosomes (Cimino-Reale et al. 2001), ce qui peut avoir des implications dans la protection des extrémités.

Organismes	Longueur d-b	Longueur s-b	t-loop	Références
Protozoaires				
<i>Tetrahymena</i>	250-300 pb	14-15 b et 20-21 b	non	(Jacob et al. 2001) (Jacob et al. 2003)
<i>Oxytricha</i>	36 pb (macronoyau)	14 b	non	(Klobutcher et al. 1981) (Pluta et al. 1982)
	3-20 kb (micronoyau)	N/D	oui	(Murti et Prescott 1999)
<i>Euplotes</i>	38 pb	16 b	non	(Klobutcher et al. 1981)
<i>Trypanosoma</i>	10-20 kb	75-225 b	oui	(Munoz-Jordan et al. 2001)
Champignons				
<i>Saccharomyces</i>	300 +/- 75 pb	12-14 b en G1	non	(Larrivée et al. 2004)
		>30 b en phase S	non	(Wellinger et al. 1993a) (Wellinger et al. 1993b)
<i>Schyzosaccharomyces</i>	300 pb	N/D	non	
Plants				
<i>Arabidopsis</i>	2-4.5 kb	> 20-30 b	non	(Riha et al. 2000) (Riha et al. 1998)
Vertébrés				
<i>Homo Sapiens</i>	2-50 kb	> 150 b	oui	(Cimino-Reale et al. 2001) (Makarov et al. 1997) (McElligott et Wellinger 1997) (Wright et al. 1997) (Griffith et al. 1999) (Stewart et al. 2003)
Autres vertébrés	2-100 kb	> 150 b	oui	(Meyne et al. 1989) (Griffith et al. 1999)

Tableau 2 : Caractéristiques des répétitions télomériques de différents organismes.

d-b = double-brin; s-b = simple-brin; t-loop = présence de t-loop à l'extrémité des chromosomes; N/D= non-déterminé à ce jour; les longueurs mentionnées sont les moyennes des tailles tirées de la littérature, et certaines sont des estimations de longueurs.

Les données concernant la taille des extensions simple-brin présentes aux deux extrémités des chromosomes semblent être controversées (i.e. longues extensions présentes sur tous les télomères (Makarov et al. 1997), extensions asymétriques aux extrémités des chromosomes (Wright et al. 1997), forte prédominance d'extensions simple-brin plus courtes que celles mesurées auparavant sur les télomères (Cimino-Reale et al. 2001)). La taille des télomères et des extensions télomériques est déterminée par de complexes mécanismes permettant un fragile équilibre télomérique. Ces mécanismes incluent des activités diverses permettant l'allongement et le raccourcissement des télomères, toujours dans l'optique de permettre et de garder un équilibre. Les extensions simple-brin G-riches sont ubiquitaires dans les cellules de mammifères et dans les différents organismes étudiés jusqu'à présent, ce qui montre leur importance pour le maintien de la structure terminale des chromosomes et leurs différentes fonctions. Leur présence devrait être observée à chacune des extrémités des chromosomes, et leur taille, chez les mammifères, devrait permettre la formation des structures protectrices observées par microscopie électronique, les t-loops (Griffith et al. 1999; Nikitina et Woodcock 2004). Chez la levure, une activité nucléase permet de générer l'extension télomérique après la réplication du brin avancé (Wellinger et al. 1996; Dionne et Wellinger 1998; Larrivée et al. 2004; revue dans (Chakhparonian et Wellinger 2003), et un procédé similaire serait à l'œuvre chez les ciliés (Jacob et al. 2003). Une activité semblable pourrait aussi être à l'œuvre dans les cellules de mammifères pour la création des extensions télomériques requises selon leur mode de réplication (Bailey et al. 2001). Par la suite, les mécanismes de protection des extrémités seraient établis (Bailey et al. 2001).

Le raccourcissement de l'extension 3' simple-brin a récemment été identifié comme l'élément déclencheur de la sénescence répllicative chez les cellules humaines (Stewart et al. 2003). En utilisant la méthode de T-OLA (Cimino-Reale et al. 2001), le groupe de R. Weinberg a montré qu'au moment de la sénescence, l'extension 3' simple-brin des différentes lignées cellulaires étudiées est raccourcie la rendant même presque indétectable (Stewart et al. 2003). Ce raccourcissement du simple-brin télomérique serait une conséquence de la division cellulaire continue effectuée par les cellules, et non pas une conséquence de la sénescence comme telle. Une perte progressive de l'extension télomérique est observée aussi dans des cellules qui contournent la sénescence à cause de l'inactivation de p53 ou de Rb (Stewart et al. 2003). Aussi, l'expression exogène de l'enzyme télomérase permet de prévenir la perte de cette extension 3' simple-brin (Stewart et al. 2003), permettant probablement le maintien d'une structure terminale adéquate pour la formation du capuchon protecteur. En accord avec ces évidences, il a aussi été montré auparavant que si l'on inhibe la protéine télomérique humaine TRF2, impliquée dans la formation des t-loops, il y a perte partielle de l'extension 3' simple-brin, déprotection des télomères et fusions de télomères bout-à bout (van Steensel et al. 1998; Karlseder et al. 2002; Karlseder et al. 1999). Il a aussi été montré par hybridation fluorescente *in situ* (FISH, fluorescence in situ hybridization) qu'il y a accumulation de courts télomères dans des fibroblastes humains avant que les cellules n'entrent en sénescence (Martens et al. 2000). D'autres évidences portent à croire que l'extension 3' simple-brin a un rôle important à jouer dans la sénescence des cellules. Tout d'abord, des PNA (peptid nucleic acid) complémentaires à l'extension 3' simple-brin peuvent inhiber la croissance cellulaire et induire l'apoptose dans des cellules humaines immortalisées

(Shammas et al. 2004). En se liant à l'extension télomérique, les PNA pourraient mimer un raccourcissement de l'extension, ou pourraient mener à une rupture de la structure télomérique ou encore pourraient empêcher des interactions entre les protéines télomériques et l'ADN télomérique, poussant ainsi les cellules à entrer en sénescence. Aussi, l'exposition de fibroblastes humains à des oligonucléotides de séquences homologues à celles de l'extension 3' télomérique mène à la sénescence de cellules en culture (Li et al. 2003). Dans ce cas, il a été proposé que la présence d'extensions simple-brin trop courtes pour permettre la formation de la structure de t-loops pousse les cellules à entrer en sénescence (Li et al. 2003). Un raccourcissement critique des télomères ou des dommages sévères à l'ADN télomérique mèneraient à l'exposition des extensions télomériques dans la cellule, ce qui serait détecté par la cellule et mènerait à des réponses cellulaires semblables à celles mises en œuvre lors de la détection de dommages à l'ADN (Li et al. 2003).

Très récemment, ces données ont été contredites par le groupe de J. W. Shay et W.E. Wright, qui ont observé que les télomères des cellules humaines maintiennent leur extension 3' simple-brin lors de la sénescence (Chai et al. 2005). Le nouvel essai biochimique (overhang protection assay) utilisé dans cette étude met l'ADN génomique purifié en présence de protéines qui se lient spécifiquement aux extensions télomériques simple-brin (la protéine gp32 du bactériophage T4 et la protéine UP1 des mammifères). Ces protéines subissent un couplage (crosslink) à l'ADN avec le glutaraldéhyde, l'ADN dépourvu de protéines est digéré avec l'ADNase I. Le couplage est par la suite renversé, les protéines sont éliminées par la protéinase K, puis les fragments d'ADN sont migrés sur un gel non-dénaturant et hybridés avec une sonde radioactive spécifique pour détecter

les séquences télomériques humaines (Chai et al. 2005). Les résultats de cette étude indiquent que les extensions télomériques des fibroblastes ne semblent pas diminuer lors de l'entrée des cellules en sénescence (Chai et al. 2005), et les auteurs suggèrent que le raccourcissement observé auparavant n'est pas la cause de la sénescence cellulaire. Les différents résultats des deux études pourraient être causés par les différentes conditions utilisées lors de la culture des cellules et ainsi conduire à des niveaux de stress ou de dommages à l'ADN qui sont différents dans les cellules (Chai et al. 2005). Il a été montré par Weinberg et collaborateurs que certaines sortes de stress peuvent conduire à une diminution de la taille des extensions télomériques, e.g. l'exposition au méthyl-nitro-nitroso-guanidine, au peroxyde d'hydrogène et à la cycloheximide (Stewart et al. 2003), produits agissant probablement sur l'ADN G-riche tels les télomères (Petersen et al. 1998). Cependant, aucune réduction dans la taille des extensions n'a été aussi forte que celle observée lors de la sénescence des cellules, suggérant que le raccourcissement des extensions ne constitue pas une réponse cellulaire au stress physiologique (Stewart et al. 2003). Une réduction de taille des extensions a été rapportée suite à la culture cellulaire continue (Keys et al. 2004), et il demeure possible que le raccourcissement des extensions observé lors de la sénescence soit une conséquence du stress accumulé par les cellules lors de la culture *in vitro* continue (Stewart et al. 2003). En se basant sur des résultats antérieurs publiés par le même groupe, les auteurs disent que pour des fibroblastes normaux, seulement 10% des télomères possédant les extensions les plus courtes seraient la cause de la sénescence (Zou et al. 2004). Il a été montré auparavant que la viabilité des cellules et la stabilité des chromosomes dépend de la taille du télomère le plus court et non pas de la taille moyenne de tous les télomères (Hemann et

al. 2001). En effet, le télomère le plus court de la cellule est celui qui devient critique pour la viabilité cellulaire et la stabilité des chromosomes (Hemann et al.2001). Si ces données sont vraies, les auteurs disent qu'il est donc impossible de détecter une diminution globale de la taille de l'extension simple-brin sur 9-10 télomères individuels qui sont comparés à 92 télomères avec des extensions normales. La perte de séquences à un seul télomère peut mener à de graves conséquences pour la cellule, mais peut s'avérer difficile à détecter par les méthodes d'analyse utilisées. De nouvelles méthodes devront être développées pour pouvoir regarder un télomère précis, suivre son évolution durant les divisions cellulaires et déterminer les causes/conséquences de la sénescence cellulaire.

Objectifs du projet

Étant donné l'importance des extensions simple-brin télomériques et le fait que nos connaissances sont partielles et parfois contradictoires à leur sujet, je me suis intéressée à développer de nouvelles techniques pour la détection et la caractérisation de ces extensions télomériques. Avant le début de mes travaux, il était connu que les télomères de levure acquièrent des extensions simple-brin G-riches tard en phase S, lors de la réplication de l'ADN télomérique (Wellinger et al. 1993b; Wellinger et al. 1993a; Marcand et al. 2000; Wellinger et al. 1996; Diede and Gottschling 1999). Cependant, la présence de ces extensions dans les autres phases du cycle cellulaire était jusqu'à présent indéterminée. De plus, les complexes protéiques impliqués dans la création et le maintien de cette extension télomérique au brin avancé lors de la réplication de l'ADN étaient également inconnus. L'hypothèse de départ était que la longueur des extensions simple-brin chez la levure en phase G1 du cycle cellulaire est plus courte qu'en phase S, ce qui explique la difficulté à la détecter et à la mesurer par la méthode traditionnelle d'hybridation « in gel ». Les objectifs du projet étaient de mettre au point une méthode permettant de caractériser les extensions télomériques simple-brin de la levure en phase G1 du cycle cellulaire, d'identifier des facteurs cellulaires impliqués dans leur création et leur maintien principalement au brin avancé, et de mieux caractériser leur action au niveau des télomères.

Par la suite, je me suis intéressée à la mise au point d'une méthode plus sensible pour la caractérisation des extensions simple-brin télomériques de cellules humaines, et je voulais en particulier mieux caractériser les extensions télomériques des cellules ALT.

De plus, avec cette méthode plus sensible, il serait possible de déterminer s'il y a un raccourcissement de l'extension simple-brin lors de l'entrée des cellules en sénescence.

Généralités concernant l'épissage et l'épissage alternatif

Chez les eucaryotes, les ARNs messagers sont d'abord transcrits en tant que précurseurs (ARN pré-messagers), contenant des séquences encodant généralement de futures protéines (exons) et des séquences dites non-codantes (introns) qui sont intercalées entre les différents exons. Ces séquences non-codantes sont par la suite enlevées, pour que les exons soient ligués ensemble pour former l'ARN messenger mature, qui sera ensuite traduit en protéines. L'épissage est un processus cellulaire important pour la régulation de l'expression des gènes, chez toutes les espèces, de la bactérie à l'humain. Grâce aux avancées permises par le séquençage complet du génome humain, certaines estimations permettent de saisir l'ampleur et les conséquences de l'épissage.

Il a été calculé qu'un gène humain moyen contient environ 8.8 exons, chacun de ces exons comptant en moyenne 145 nucléotides. La taille moyenne d'un intron est de 3365 nucléotides, et les régions 5' UTR (untranslated regions) et 3' UTR comptent 770 et 300 nucléotides en moyenne, respectivement. Ceci mène à un grand total, pour un gène « standard », d'environ 27 kb. Après les différentes étapes de maturation de l'ARN pré-messenger, l'ARN messenger moyen qui est exporté au cytosol est constitué de 1340 nucléotides de séquence codante, 1070 nucléotides de régions non-traduites et une queue poly-A (Lander et al. 2001). Ceci démontre que plus de 90% de l'ARN pré-messenger est constitué d'introns qui sont enlevés, et que seulement 10% de l'ARN pré-messenger moyen est composé de séquences exoniques qui sont épissées pour former l'ARN messenger (Lander et al. 2001; Stamm et al. 2005). L'épissage est donc un mécanisme qui doit être extrêmement bien régulé pour ne pas mener à des conséquences catastrophiques pour la cellule.

Dans un processus cellulaire étroitement relié à l'épissage et appelé épissage alternatif, un exon peut être incorporé dans l'ARN messager mature ou il peut être excisé comme un intron, selon différentes conditions. Initialement, il a été pensé que le processus d'épissage alternatif était plutôt l'exception que la règle, et que seulement 5% des gènes seraient affectés par l'épissage alternatif (Sharp 1994). Maintenant, il a été montré que la majorité des gènes est sujette à l'épissage alternatif. En fait, avec le projet de séquençage complet du génome humain, il a été constaté que le génome humain contenait beaucoup moins de gènes que ce qui avait été estimé au départ (Lander et al. 2001; Pennisi 2000). C'est alors que l'étendue d'action de l'épissage alternatif a commencé à être mieux comprise (Lareau et al. 2004), car il est un moyen utilisé par de nombreux organismes afin de générer une grande diversité dans leur contenu en protéines (Graveley 2001). En fait, ce mécanisme serait responsable du fait que seulement 23 000-40 000 gènes humains (Lander et al. 2001; Venter et al. 2001) permettent de créer un protéome (ensemble de protéines contenues dans un génome) de plus de 90 000 protéines (Harrison et al. 2002). L'ancienne hypothèse qu'un gène produit une protéine ne peut expliquer la complexité et la diversité du contenu en protéines des différents organismes vivants, donc toutes ces données portent à croire que l'épissage alternatif est finalement la règle, et non pas l'exception.

Certaines données sur l'épissage alternatif de deux chromosomes humains, soit les chromosomes 22 et 19 sont connues (Lander et al. 2001). Des 245 gènes présents sur le chromosome 22, 59% sont épissés de façon alternative, et les 544 gènes présents sur le chromosome 19 donnent naissance à 1859 ARNs messagers différents (Stamm 2002). Des cas d'épissage alternatif extrême existent : par exemple, le gène humain de neurexine

3 peut potentiellement mener à 1 728 transcrits grâce à l'épissage alternatif (Stamm et al. 2005). Aussi, chez la drosophile, le gène DSCAM (Down sndrome cell adhesion molecule) peut potentiellement générer 38 016 isoformes différents grâce à l'épissage alternatif (Celotto and Graveley 2001; Stamm et al. 2005). Dans certains cas, l'épissage alternatif peut permettre de générer plus de transcrits à partir d'un simple gène que le nombre total de gènes inclus dans un génome entier (Schmucker et al. 2000; Graveley 2001; Stamm et al. 2005).

Une grande partie des événements d'épissage alternatif sont régulés de façon spécifique principalement selon le type de cellule, le stade développemental et les réponses à des stimuli externes (Faustino et Cooper 2003). Aussi, étant donné l'immense précision requise pour l'enlèvement des introns, la complexité du processus d'épissage alternatif et son importance dans la régulation de l'expression des gènes dans la cellule, il n'est pas surprenant de constater que des failles à plusieurs niveaux dans ce processus peuvent éventuellement mener à des maladies génétiques (Faustino et Cooper 2003; Boue et al. 2003; Stamm et al. 2005). En effet, une étude effectuée à ce sujet il y a plus de 12 ans montre qu'environ 15% des mutations ponctuelles qui donnent lieu à des maladies génétiques chez l'humain, sont dues à la destruction des sites d'épissage (Krawczak et al. 1992). De nos jours, il est évident que ce résultat était une sous-estimation, car il tenait seulement compte des mutations dans les séquences des sites d'épissage classiques, les seuls connus à cette époque. Le mécanisme d'épissage est également affecté par tous les types de mutations pouvant altérer les éléments de régulation de l'épissage (Faustino et Cooper 2003), par des mutations qui empêchent la production de protéines fonctionnelles (Cartegni et al. 2002) et par des mutations dans des sites d'épissage qui affectent la

concentration nucléaire de plusieurs protéines, ce qui peut mener à différentes maladies. Tous ces changements dans le patron d'épissage alternatif peuvent être la cause et la conséquence de pathologies humaines (Philips et Cooper 2000; Stamm 2002).

Tous les eucaryotes possèdent des introns dans au moins certains de leurs gènes, mais le nombre peut varier considérablement d'un organisme à un autre. Par exemple, la grande majorité des ~ 35 000 gènes humains contient des introns (Ewing et Green 2000; Roest Crollius et al. 2000), alors que pour la levure *S. cerevisiae*, seulement 250 des 6000 gènes contiennent des introns (Lopez et Seraphin 2000). De ces 250 gènes de levure contenant des introns, seulement un très faible nombre est connu pour encoder des transcrits qui sont épissés de façon alternative (Davis et al. 2000; Spingola et Ares 2000). Des études comparatives entre l'humain et la souris ont été effectuées et il en ressort que le taux de conservation des événements d'épissage alternatif est variable entre les mammifères (Modrek et Lee 2003). Cependant, des comparaisons de banques de données de ESTs indiquent que des niveaux similaires d'épissage alternatif ont lieu dans des espèces distinctes comme l'humain, la souris, la drosophile ou le ver *C. elegans*, ce qui met l'accent sur l'importance de l'épissage alternatif à travers l'évolution (Brett et al. 2002; Stamm et al. 2005).

Les mécanismes et la machinerie de base de l'épissage



Figure 9. Multiples signaux reconnus par les facteurs d'épissage lors de la formation du spliceosome. Les signaux d'épissage classiques, tels que retrouvés sur environ 99% des introns permettent la reconnaissance de tous les exons. Il peut aussi y avoir d'autres signaux présents dans une classe mineure d'introns, et reconnus par une forme alternative du spliceosome (Tarn et Steitz 1997). Tiré de Faustino et Cooper (2003).

L'épissage est un processus partagé

par tous les eucaryotes qui permet la formation des ARNs messagers à partir de précurseurs beaucoup plus longs, et s'avère être un excellent moyen de régulation génétique (Black 2003). Le

but de l'épissage est simple et évident : enlever les séquences introniques des transcrits primaires d'ARN pré-messager pour mener à la formation de l'ARN mature qui pourra être traduit par la suite. De subtils changements dans le choix des sites d'épissage peuvent avoir des répercussions très variables sur la production de l'ARN messager et des protéines. Pour aider à la précision du processus d'épissage, certaines séquences conservées sont retrouvées aux différentes jonctions des introns/exons, et sont appelées sites d'épissage. Le site d'épissage 5' (5' splice site) est situé à la jonction de l'intron et de l'exon, plus spécifiquement à l'extrémité 5' de l'intron. Ce site d'épissage comprend un dinucléotide GU à l'extrémité 5' de l'intron, ainsi qu'une séquence consensus qui peut être moins bien conservée (figure 9) (revue dans (Black 2003)). À l'autre extrémité de l'intron, le site d'épissage 3' (3' splice site) possède 3 éléments conservés. Tout d'abord, le site de branchement (branch site). Celui-ci est suivi d'une région riche en pyrimidines (polypyrimidine tract ou Y tract) qui est située généralement de 20 à 40 nucléotides en aval du site d'épissage 3'. Suit par après le site d'épissage 3' comme tel, qui possède un dinucléotide AG à l'extrémité 3' de l'intron (figure 9). Plusieurs détails concernant les procédures de reconnaissance des introns par le spliceosome sont encore mal compris, et sont le sujet de recherches continues.

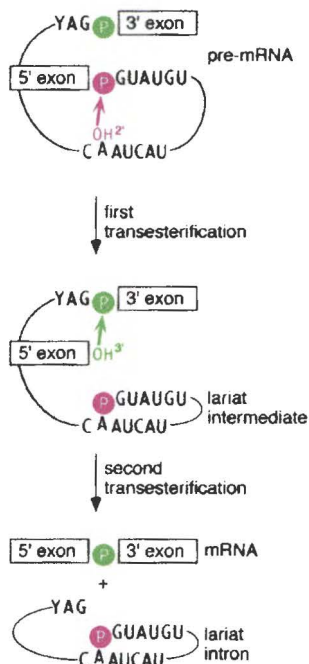


Figure 10. Réactions de trans-estérification menant à l'épissage des ARNs pré-messagers. Voir le texte pour les détails. Tiré de Staley and Guthrie (1998), revue dans Black (2003).

La réaction d'épissage peut être séparée en deux étapes biochimiques distinctes (figure 10) (Staley et Guthrie 1998; Nilsen 2003; Black 2003). Ce sont deux réactions consécutives de trans-estérification, qui remplacent un lien phospho-diester par un autre. D'abord, il y a une attaque nucléophile du groupement 2' hydroxyl (2' OH) de l'adénosine du site de branchement sur le site d'épissage 5'. Cette attaque cause l'hydrolyse du lien phospho-diester au site d'épissage 5', et mène à la formation d'une liaison entre le site de branchement et le nucléotide de l'intron en position 5' terminale qui a été relâché. Les produits de cette première réaction sont donc un exon libéré et un intron sous la forme de lasso (lariat) qui est lié à l'exon 3'. Par la suite, l'extrémité 3' hydroxyl (3' OH) libre de

l'exon 5' relâché lors de la première réaction, attaque la liaison phospho-diester du site d'épissage 3'. Il y a bris du lien phospho-diester à cet endroit et formation d'une nouvelle liaison phospho-diester entre l'exon 5' et l'exon 3'. Ainsi, les produits finaux relâchés par la réaction d'épissage complétée sont les exons liés entre eux et l'intron sous forme de lasso.

Chez les eucaryotes, le spliceosome constitue la machinerie cellulaire responsable de l'épissage des introns. Le spliceosome majeur est responsable de l'épissage de la grande majorité des introns contenus dans les ARNs pré-messagers, appelé introns de type U2. Il est formé par l'assemblage de 5 ribonucléoparticules, appelées snRNP U1, U2, U4/U6·U5 (small nuclear ribonucleoparticule). Chacune de ces particules contient

une petite composante ARN stable (snRNA, small nuclear RNA) qui est liée par plusieurs protéines et facteurs d'épissage (Nilsen 2002; Will et Lührmann 2001; Staley et Guthrie 1998). Il existe aussi un spliceosome mineur beaucoup moins abondant qui est responsable de l'épissage d'une classe spéciale d'introns nommés introns de classe U12. Ce spliceosome mineur est composé des snRNPs U11, U12, U4atac, U6atac et U5 (Tarn et Steitz 1997; Will et Lührmann 2001; Brow 2002; Patel et Steitz 2003), et il reconnaît des signaux différents de ceux reconnus par le spliceosome majeur. Il a été récemment montré que le spliceosome est en fait un méga-complexe ribonucléoprotéique comprenant environ 300 protéines différentes (Rappsilber et al. 2002; Zhou et al. 2002; revue dans Nilsen 2003; Jurica et Moore 2003)). Différents résultats concernant la composition protéique et la purification des spliceosomes de mammifères ont été obtenus par divers groupes et sont dépendants de l'efficacité des méthodes de purification et d'identification des protéines, qui évoluent rapidement (revue dans Jurica et Moore 2003). Chacune des différentes protéines retrouvées dans le spliceosome doit avoir un rôle à jouer dans l'épissage des ARNs pré-messagers.

Le spliceosome doit permettre de joindre de façon très précise et efficace deux sites d'épissage souvent séparés par plusieurs kilobases. Il doit reconnaître les bons sites d'épissage parmi de nombreux sites d'épissage potentiels (sites cryptiques) dispersés dans les introns, ce qui implique des mécanismes complexes afin de les identifier correctement. En ce qui concerne l'épissage alternatif, les mêmes signaux sont reconnus par le spliceosome, mais en plus, il y a des éléments de modulation spécifiques pour l'utilisation ou le blocage de certains sites (voir plus loin). Les étapes menant à la

formation du spliceosome actif et les mécanismes permettant la sélection des sites d'épissage aident à mieux comprendre la modulation de l'épissage.

L'identification des sites d'épissage 5' et 3' constitue la première étape d'assemblage du spliceosome, (revue dans Jurica et Moore 2003; Nilsen 2003). D'abord, le snRNP U1 reconnaît le site d'épissage 5' en grande partie grâce à des interactions entre sa propre composante ARN (snRNA) et l'ARN pré-messager. Ces interactions peuvent être facilitées et stabilisées par de nombreuses autres protéines et/ou facteurs auxiliaires. Les éléments du site d'épissage 3' sont aussi liés par des protéines spécifiques. SF1 est une protéine qui lie le site de branchement (cette protéine est appelée BBP chez la levure, branchpoint binding protein). Le facteur d'épissage U2AF (U2 auxiliary factor) reconnaît le site d'épissage 3'. Ce facteur d'épissage est composé de deux sous-unités (U2AF³⁵ et U2AF⁶⁵, pour les sous-unités de 35 kDa et de 65 kDa, respectivement) et chacune possède un domaine RS (Zamore et Green 1989; Zhang et al. 1992; Zamore et

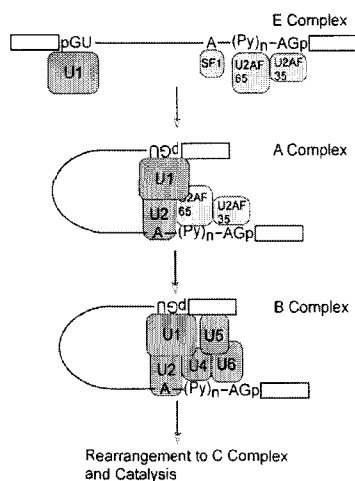


Figure 11. Formation du spliceosome et des différents complexes sur l'intron. Voir la description détaillée des événements dans le texte. Tiré de Black (2003) et revue dans Jurica et Moore (2003).

al. 1992). La liaison de U2AF³⁵ a lieu à la jonction du site d'épissage 3', qui comprend deux résidus invariables AG au site de liaison spécifique. La sous-unité U2AF⁶⁵ se lie à la région riche en pyrimidines, positionnée près du site de branchement (Zamore et al. 1992). Le complexe ainsi formé constitue le plus précoce dans l'assemblage du spliceosome, appelé complexe E (early) ou complexe d'engagement (commitment complex). Ce complexe contient snRNP U1 et U2AF liés aux deux extrémités de l'intron (Das et al. 2000; revue dans Black 2003). Ce complexe est par la suite

rejoint par le snRNP U2 qui se lie au site de branchement, via des interactions ARN : ARN entre sa composante snRNA et l'ARN pré-messager (Maniatis et Reed 2002), pour former le complexe A. Le tri-snRNP U4/U6/U5 rejoint par la suite le complexe A pour donner naissance au complexe B. Ce complexe subit un réarrangement exhaustif pour former le complexe C, dans lequel l'interaction du snRNP U1 au site d'épissage 5' est remplacée par le snRNP U6, puis il y a perte séquentielle des snRNPs U1 et U4 de ce complexe. C'est le complexe C qui permet la catalyse des deux réactions menant à l'épissage. Suite à la réaction d'épissage, les différentes composantes du spliceosome seraient recyclées pour permettre d'autres rondes d'épissage subséquentes. Cependant, certaines évidences récentes obtenues chez la levure permettent de croire que le spliceosome existerait sous une forme pré-assemblée comprenant plusieurs sous-unités (Stevens et al. 2002). De plus, de nombreuses évidences montrent maintenant que tous les processus reliés à la production de l'ARN (e.g. transcription, maturation de l'ARN, exportation de l'ARN mature au cytoplasme, mécanismes de surveillance) sont reliés entre eux, et seraient accomplis dans ce que l'on peut appeler des « usines complexes de l'expression génétique » (revue dans Maniatis et Reed 2002). Ainsi, chaque étape de l'expression d'un gène, de la plus précoce à la plus tardive, serait reliée à ce grand réseau de production, et l'épissage en est une portion intégrante.

Dans la plupart des systèmes étudiés, les changements dans la sélection des sites d'épissage sont régulés par des modifications dans la liaison des facteurs du spliceosome. À cause de la dégénérescence des séquences consensus présentes aux sites d'épissage, la formation des différents complexes menant à un spliceosome actif doit être modulée par des interactions avec de nombreux autres facteurs protéiques (Lallena et al. 2002; Berget

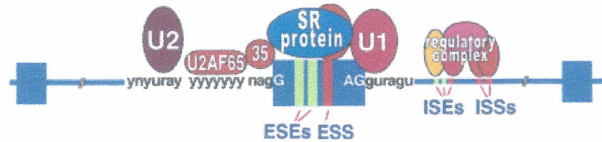


Figure 12. Facteurs et sites d'épissage permettant la définition de l'exon. Les facteurs d'épissage classiques se lient aux sites de liaison spécifiques 5' et 3' permettant la définition de l'exon. Ce processus peut être aidé par des interactions avec des protéines spécifiques. Les facteurs se lient à des sites auxiliaires (ESEs, ESSs) dont les protéines SR sont généralement requis pour l'épissage constitutif et l'épissage alternatif des exons. Les éléments introniques (ISEs, ISSs) servent aussi à moduler l'épissage alternatif de façon spécifique. Tiré de Faustino (2003).

1995). Il a été remarqué que la présence d'un site 5' stimule l'excision d'un intron placé en amont (Robberson et al. 1990; Hoffman et Grabowski 1992; Berget 1995). En effet, les différentes composantes du spliceosome, liées de part et d'autre d'un exon en 5' et en 3', peuvent interagir entre elles pour stimuler l'excision des introns, un

processus nommé « définition de l'exon » (exon definition) (figure 12) (Hoffman et Grabowski 1992; Robberson et al. 1990). L'exon est donc compris entre un site de liaison pour U1 et un site de liaison pour U2AF. Il a été proposé que la liaison de U1 à un site 5' stimulerait la liaison de U2AF à un site 3' en amont (Robberson et al. 1990), permettant la définition de l'exon. Les interactions des facteurs d'épissage sont également aidées et stabilisées par des interactions protéine : protéine, dont plusieurs font partie de la famille des protéines SR (voir plus loin). En plus du processus aidant à la définition de l'exon, il y a de nombreuses autres séquences régulatrices qui affectent l'assemblage du spliceosome. En fait, des éléments qui agissent pour stimuler l'assemblage du spliceosome peuvent être retrouvés dans les exons ou dans les introns (ESEs, exonic splicing enhancers ou ISEs, intronic splicing enhancers). Vice-versa, des éléments bloquant l'assemblage du spliceosome ou l'utilisation de sites d'épissage peuvent également être retrouvés dans l'ARN (ESS, exonic splicing silencers ou ISSs, intronic

splicing silenciers). Les mécanismes d'action de ces différents éléments peuvent être très complexes, et certains seront abordés dans les sections suivantes.

Les protéines SR constituent une grande famille de protéines se liant à des séquences spécifiques retrouvées dans l'ARN et permettant l'activation de l'épissage via différents moyens (Graveley 2000; Manley et Tacke 1996; Cartegni et al. 2002). Les multiples membres de la famille des protéines SR ont une structure commune qui comporte un ou deux domaines de liaison à l'ARN (RRM, RNA recognition motif ou KH domain, K-homology domain), suivis par un domaine RS contenant des répétitions arginine/sérine (Graveley 2000). Les deux domaines ont des fonctions particulières. Les domaines RRM permettent la liaison à l'ARN, ciblant ainsi une protéine à un élément exonique particulier (Chandler et al. 1997; Mayeda et al. 1999; Tacke and Manley 1995; Lavigne et al. 1993), alors que les domaines RS, largement interchangeables, permettent de stimuler l'épissage (Chandler et al. 1997; Wang et al. 1998). Les résidus sérine du domaine RS peuvent être phosphorylés par différentes kinases, ce qui peut modifier leur activité dans l'épissage (Sacco-Bubulya et Spector 2002; Sanford et Bruzik 2001). Cette phosphorylation est requise pour l'activité des protéines dans l'épissage (Prasad et al. 1999; Sanford et Bruzik 1999; Mermoud et al. 1994), et les domaines RS phosphorylés pourraient permettre des interactions protéine : protéine (Xiao et Manley 1997).

Les protéines SR semblent avoir des rôles distincts dans l'épissage constitutif et l'épissage alternatif. L'épissage dans des extraits cellulaires requiert la présence d'au moins une protéine SR, et à ce niveau la plupart des membres de la famille sont interchangeables, mais elles ne sont pas interchangeables *in vivo* (Longman et al. 2000;

Wang et al. 2001). Une des activités associées à ces protéines est l'activation de l'épissage par la liaison des séquences exoniques stimulant l'épissage (exonic splicing enhancer, ESEs). Plusieurs exons contiennent en effet des éléments ESEs auxquels se lient des membres de la famille des protéines SR (Cartegni et al. 2002; Blencowe 2000), qui interagissent aussi avec les composantes du spliceosome ce qui semble être un mécanisme général pour la définition de l'exon (figure 12). À ce sujet, il a été montré que la protéine ASF/SF2, membre de la famille des protéines SR, interagit avec le domaine RS du snRNP U1, U1 70K (Kohtz et al. 1994). En opposition aux ESEs liés par des protéines SR, il y a les séquences ESSs qui, également par la liaison de facteurs spécifiques, répriment l'utilisation de certains sites d'épissage. Ces éléments sont souvent liés par les protéines hnRNPs (heterogeneous nuclear ribonucleoproteins) (voir plus loin).

L'épissage alternatif

L'épissage alternatif est le mécanisme qui permet l'association de différents sites d'épissage 5' et 3' à partir d'un ARN messager. Ce procédé permet d'inclure ou d'exclure certains exons afin de produire de multiples ARNs messagers différents. Ultimement, ce procédé mène à la formation de plusieurs isoformes protéiques à partir d'un seul et unique gène. L'épissage alternatif est un contributeur majeur de la diversité protéique observée dans plusieurs organismes, autant les métazoaires que les mammifères (Maniatis et Tasic 2002; Stamm et al. 2005). L'épissage est aussi un moyen de moduler l'expression de certains gènes de façon spatio-temporelle ou en réponse à différents stimuli (revue dans Stamm 2002). Par exemple, l'expression de certains isoformes des protéines impliquées dans le développement du système nerveux est régulée de façon très

précise durant le développement et les différents types cellulaires (revue dans Stamm 2002; Daoud et al. 1999; Vallano et al. 1996).

La plupart des concepts connus aujourd'hui concernant la régulation de l'épissage alternatif viennent des études faites chez la drosophile. Chez cet organisme, la détermination du sexe de la mouche est le résultat d'une cascade d'événements modulés au niveau de la transcription et de l'épissage alternatif (revue dans Chabot 1996; Adams et al. 1997; Black 2003). Il a été montré que le gène principal de la cascade encode la protéine Sex lethal (Sxl), une protéine liant l'ARN et présente uniquement chez les mouches femelles. L'expression de la protéine Sxl permet la répression des patrons d'épissage menant à des mouches mâles (Baker 1989; revue dans Black 2003). Une des cibles de la protéine Sxl est le transcrit du gène *transformer* (Tra). Chez les mouches mâles, en absence du facteur Sxl, un site d'épissage 3' fort présent dans l'ARN pré-

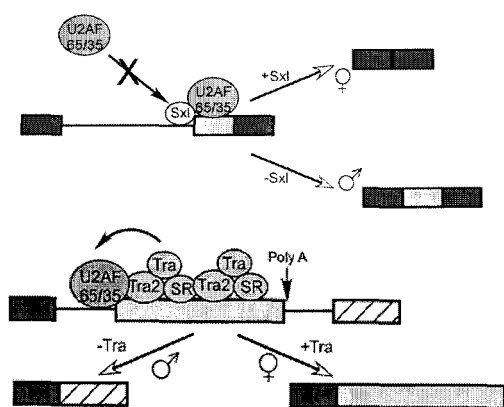


Figure 13. Régulation de l'épissage chez la drosophile. A) La protéine Sxl uniquement présente chez les femelles, bloque l'utilisation d'un site d'épissage 3' et mène à l'expression d'une protéine Tra fonctionnelle. B) Régulation du transcrit Doublesex. La protéine Tra produite chez les femelles est un régulateur positif de l'épissage de Doublesex (Dsx), encodant deux facteurs différents. Dsx est exprimée chez les femelles et ainsi bloque la différenciation en mâles. Tiré de Black (2003).

messenger de Tra est utilisé. L'ARN messenger ainsi obtenu contient un arrêt de transcription précoce et mène à la production d'une protéine Tra tronquée et non-fonctionnelle. Chez les mouches femelles, l'expression de la protéine Sxl est dépendante de l'activation de la transcription à partir d'un promoteur sensible à la présence de deux chromosomes X. Le facteur Sxl lie l'ARN pré-messenger encodant la

protéine Tra et bloque la liaison de U2AF au site d'épissage mâle-spécifique, déplaçant ainsi sa liaison à un site d'épissage 3' qui est femelle-spécifique (Valcarcel et al. 1993). L'utilisation de ce site d'épissage mène à la production d'une protéine Tra complète et active, chez les mouches femelles (figure 13A). Cette protéine Tra est un régulateur positif de l'épissage. Elle active des patrons d'épissage spécifiques aux mouches femelles en se liant à ses cibles, dont le transcrit Doublesex (Dsx). Donc, la protéine Tra, la protéine Tra2 et la protéine SR RBP1 permettent d'assembler un complexe stimulateur d'épissage, sur une séquence présente dans un exon alternatif de l'ARN pré-messager de Dsx. La formation de ce complexe multiprotéique aide au recrutement du facteur U2AF⁶⁵ au site d'épissage 3' faible, et mène à la production d'un régulateur transcriptionnel qui permet la différenciation en femelle. Chez les mouches mâles, la protéine Tra est absente, et l'ARN messager de Dsx est épissé pour mener à la production d'un facteur de Soma

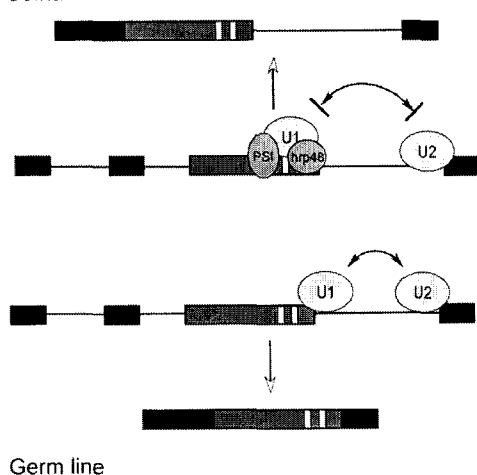


Figure 14. Régulation de l'épissage de l'élément P chez la drosophile. Dans les cellules somatiques, l'expression du facteur PSI mène à la formation d'un complexe avec les protéines hnp48 et snRNP U1 à un pseudo-site d'épissage 5' et empêche l'utilisation du réel site d'épissage 5'. Tiré de Black (2003).

transcription qui détermine la différenciation en mâle (figure 13B) (Coschigano et Wensink 1993; Tian et Maniatis 1993; Hertel et al. 1996; Lynch et Maniatis 1996).

Un autre exemple d'épissage alternatif bien caractérisé chez la drosophile est le gène codant pour la transposase P (figure 14) (revue dans Adams et al. 1997; Chabot 1996; Black 2003). La transposition des éléments P a lieu seulement dans les cellules germinales de la

mouche. Dans les cellules somatiques, il n'y a pas de transposition à cause de l'inhibition de l'expression de la transposase. En effet, dans les cellules somatiques de la mouche, le troisième intron de l'élément P est retenu dans l'ARN messager, prévenant la traduction de l'enzyme transposase et bloquant ainsi la transposition somatique (Laski et al. 1986). Il a été montré que l'exon en amont du 3^e intron contient une séquence de régulation qui inhibe l'épissage dans les cellules somatiques (Siebel et al. 1992). Cette région contient des sites de liaison pour les protéines PSI (P element Somatic Inhibitor) et la protéine hrp48, qui est un homologue chez la drosophile des protéines hnRNPs A et B retrouvées chez les mammifères (Siebel et al. 1992; Siebel et al. 1995; Adams et al. 1997). La région C-terminale de PSI peut lier la protéine U1 70K (Labourier et al. 2001). Donc, dans les cellules somatiques, PSI se lie à l'élément répresseur retrouvé dans l'exon, et par son interaction avec la protéine U1 70K, induit la liaison du snRNP U1 à un pseudo-site d'épissage 5' présent dans l'élément, mais généralement non-utilisé pour l'épissage (Siebel et al. 1992). Ceci prévient donc l'utilisation du site d'épissage 5' normal, qui par contre est utilisé dans les cellules germinales.

Ces observations, et plusieurs autres, ont donc permis d'identifier différents éléments modulant l'épissage alternatif. Certains éléments sont inhibiteurs, certains éléments sont stimulateurs, et on les retrouve à la fois dans les exons et dans les introns. De plus, ces études ont permis de mieux comprendre les mécanismes d'action de différents éléments; certains éléments modulent l'épissage par la formation de structures secondaires tandis que d'autres nécessitent la liaison de facteurs protéiques sur l'ARN pour permettre leur action.

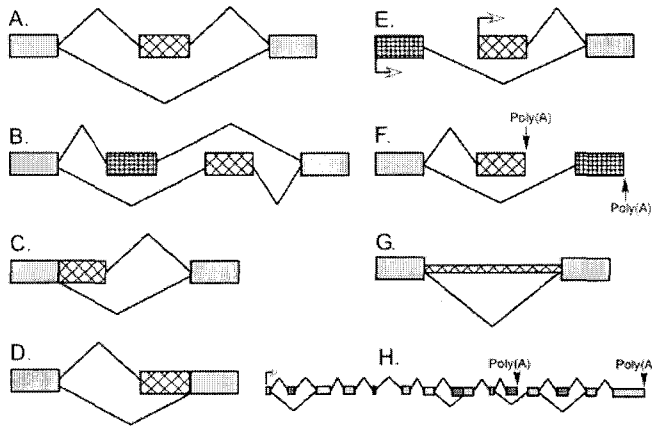


Figure 15. Différents patrons possibles pour l'épissage alternatif. Voir le texte pour les détails. Tiré de (Black 2003).

Les résultats obtenus chez la drosophile et de nombreux autres obtenus chez d'autres espèces ont permis de conclure que de multiples patrons différents d'épissage existent dans la nature (revue dans Black 2003; Graveley 2001; Chabot 1996)). Certains exons peuvent être parfois inclus, parfois exclus, et sont

appelés des exons cassettes (figure 15A). Dans certains cas, des exons cassettes peuvent être mutuellement exclusifs, menant à la production d'ARN messagers incluant toujours un seul des nombreux choix d'exons possibles (figure 15B). Les exons peuvent aussi posséder de multiples sites d'épissage 5' ou 3' qui sont utilisés de façon alternative (figures 15C, 15D). Certains exons peuvent utiliser des promoteurs alternatifs et ainsi changer leur exon 5' terminal, ou encore utiliser des sites de polyadénylation alternatifs ce qui permet de changer leur extrémité 3' (figures 15E, 15F). Il peut aussi y avoir des cas où des séquences introniques sont retenues dans la séquence finale de l'ARN messager (figure 15G). De plus, un même transcrit peut contenir plusieurs différents sites d'épissage alternatif (figure 15H). Ces sites sont souvent utilisés de manière combinée, pour produire différents isoformes d'ARN messagers. Les signaux d'épissage retrouvés dans les exons alternatifs sont généralement faibles et dérivés des séquences consensus, ce qui peut permettre une plus grande flexibilité de sites à utiliser.

La présence d'éléments structuraux dans l'ARN même peut moduler la sélection des sites d'épissage. Chez les mammifères, la première démonstration d'éléments structuraux capables d'affecter l'épissage a été faite par insertion de séquences formant des duplexes stables, dans des ARNs servant de modèles (Solnick 1985). Par la suite, il a été montré qu'il est possible de moduler l'épissage alternatif en masquant un site d'épissage par la formation d'un duplexe intramoléculaire ou en permettant de rapprocher deux sites d'épissage qui étaient distants dans l'espace (Solnick 1985; Eperon et al. 1986; Solnick et Lee 1987). De nombreux éléments naturels permettant la formation de duplexes et la modulation de l'épissage alternatif ont aussi été mis en évidence dans certains gènes (Clouet d'Orval et al. 1991a; Clouet d'Orval et al. 1991b; Libri et al. 1991; Blanchette et Chabot 1997; Côté et Chabot 1997; Muro et al. 1998). Dans certains cas, ces structures bloquent l'accès du spliceosome aux sites d'épissage ou encore déstabilisent les liaisons des facteurs aux sites d'épissage. Les différents éléments structuraux ont un rôle mineur à jouer dans l'épissage alternatif. La majorité des éléments de régulation identifiés jusqu'à maintenant impliquent surtout des facteurs nucléaires, et c'est la liaison de ces différents facteurs à des séquences régulatrices qui permet de stimuler ou inhiber l'utilisation de sites d'épissage. L'épissage alternatif étant un mécanisme complexe, sa régulation est souvent le résultat d'un équilibre très précis entre de multiples éléments positifs et négatifs. De plus, étant donné que l'épissage est relié à tous les autres processus de production de l'ARN messager, il est certain que de nombreux autres procédés et leurs étapes de contrôle influencent aussi l'épissage (Maniatis et Reed 2002).

Les protéines hnRNP A1 et hnRNP A2

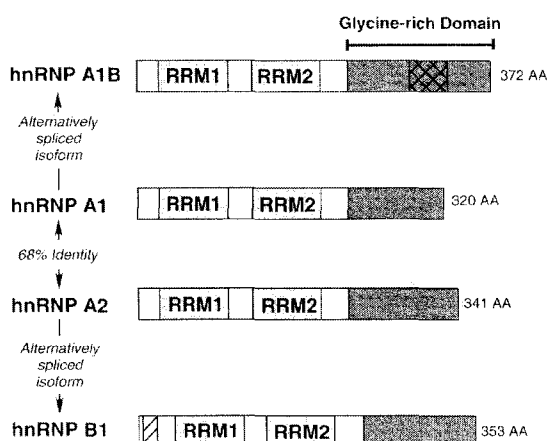


Figure 16. Famille des protéines hnRNP A1/A2. Les domaines de liaison à l'ARN (RRM) et le domaine riche en glycine sont indiqués. Les différences entre A1 et A1^B, entre A2 et B1 sont indiqués par la portion hachurée. Tiré de Chabot et al. (2003).

Les protéines hnRNPs ont été décrites il y a quelques années comme un groupe de protéines liant l'ARN associé à la chromatine (chromatin-associated RNA-binding protein) (Dreyfuss et al. 1993; Krecic and Swanson 1999). Suite à la transcription par l'ARN polymérase II, il y a empaquetage des ARNs pré-messagers (hnRNAs) dans des complexes ribonucléoprotéiques contenant des

particules hnRNPs (Krecic et Swanson 1999). Le noyau de base des hnRNPs est constitué de six protéines (hnRNPs A1, A2, B1, B2, C1 et C2), qui sont les plus abondantes et fortement exprimées dans des cellules en division (Beyer et al. 1977). Jusqu'à maintenant, plus de 25 de ces abondantes protéines hnRNPs ont été identifiées (Pinol-Roma et al. 1988; Dreyfuss et al. 1993; Krecic and Swanson 1999). Suite à l'immunopurification des complexes hnRNPs, l'identification des protéines associées, le clonage des gènes correspondants et des analyses structurales et fonctionnelles, il a été montré que chaque protéine hnRNP contient au moins un motif de liaison à l'ARN. De plus, il a été montré que ces protéines hnRNPs ne se lient pas de façon aléatoire à n'importe quelle séquence ARN, mais sont plutôt concentrées en des endroits précis sur l'ARN, suggérant la présence de sites de liaison optimaux (Pinol-Roma et al. 1989; Matunis et al. 1993; Wurtz et al. 1996). À cet égard, l'utilisation de la technique SELEX

a permis d'identifier de tels sites pour certaines protéines hnRNPs (Burd et Dreyfuss 1994; Gorlach et al. 1994a; Gorlach et al. 1994b). Les protéines hnRNPs ont été impliquées dans plusieurs aspects de l'expression génétique, dont la transcription, l'épissage, la polyadénylation, le transport et la stabilité des ARNs messagers.

Une des protéines hnRNP les mieux caractérisées est hnRNP A1. Elle fait partie d'un groupe de protéines ayant des activités et une structure similaires, comprenant les hnRNP A1, A1^B, A2 et B1 (figure 16) (Mayeda et al. 1994). Cette protéine possède de multiples fonctions dans la cellule (revue dans Chabot et al. 2003). Elle a entre autres été impliquée dans la stabilité (Hamilton et al. 1993; Hamilton et al. 1997) et le transport des ARNs messagers (Michael et al. 1995; Izaurralde et al. 1997), dans la liaison aux acides nucléiques (Merrill et al. 1988; Cobianchi et al. 1988; Kumar et Wilson 1990; Buvoli et al. 1990b; Munroe et Dong 1992; Burd et Dreyfuss 1994; Shamoo et al. 1994; Abdul-Manan et Williams 1996; Abdul-Manan et al. 1996), dans des interactions protéine : protéine (Cartegni et al. 1996; Swanson et Dreyfuss 1988; Buvoli et al. 1992), dans la biogénèse des télomères (LaBranche et al. 1998), ainsi que dans de la modulation de la sélection des sites d'épissage (Mayeda et Krainer 1992; Mayeda et al. 1993; Mayeda et al. 1994; Caceres et al. 1994; Yang et al. 1994; Chabot et al. 1997; Jiang et al. 1998; Bai et al. 1999; Blanchette et Chabot 1999; Hutchison et al. 2002; Nasim et al. 2002).

HnRNP A1 est une protéine nucléaire abondante de 320 acides aminés (34 kDa). Le gène encodant la protéine hnRNP A1 possède 11 exons, avec un exon cassette qui est épissé de façon alternative (exon 7B). Lorsque celui-ci est exclu de l'ARN messager il y a production de la protéine hnRNP A1, lorsqu'il est inclus dans l'ARN messager, il y a production de l'isoforme hnRNP A1^B (Buvoli et al. 1988). L'insertion de l'exon 7B dans

la région riche en glycine ajoute 52 acides aminés à la protéine (Buvoli et al. 1990a). Dans la plupart des types cellulaires, l'expression de la forme A1 est supérieure à celle de la forme A1^B (Buvoli et al. 1990a; Hanamura et al. 1998).

La protéine A1 peut être séparée en domaines (figure 16) ; le domaine amino-terminal allant jusqu'à la position 196 (aussi appelé UP1) (Williams et al. 1985a; Williams et al. 1985b; Riva et al. 1986; Merrill et al. 1987), contient deux motifs RBDs (RNA binding domains) contenant chacun deux motifs RRM, et un domaine carboxy-terminal riche en glycine (environ 40%) et arginine (GRD, Glycine-rich domain) (Casas-Finet et al. 1993). La portion amino-terminale est responsable de la liaison aux acides nucléiques (Merrill et al. 1988; Cobianchi et al. 1988; Kumar and Wilson 1990; Buvoli et al. 1990b; Munroe et Dong 1992; Shamoo et al. 1994; Abdul-Manan et Williams 1996; Abdul-Manan et al. 1996). Le site de haute affinité pour la liaison de hnRNP A1 a été identifié par SELEX (UAGGG^A/U) (Burd et Dreyfuss 1994). La portion carboxy-terminale de la protéine est responsable des interactions protéine : protéine (Cartegni et al. 1996).

Les études sur hnRNP A1 commencées dans le laboratoire étaient basées sur l'hypothèse que A1, comme certains gènes connus de la drosophile (Bell et al. 1988; Bell et al. 1991), pourrait possiblement moduler l'épissage alternatif de son propre ARN messenger. La comparaison des séquences flanquant l'exon alternatif 7B entre la souris et l'humain (figure 17), a montré un haut pourcentage de conservation des séquences introniques (60-70%), en comparaison avec seulement 25% de conservation pour les séquences introniques situées entre les exons 6 et 7 (Chabot et al. 1997), revue dans (Chabot et al. 2003). Les études ont permis de mettre en évidence 10 régions conservées,

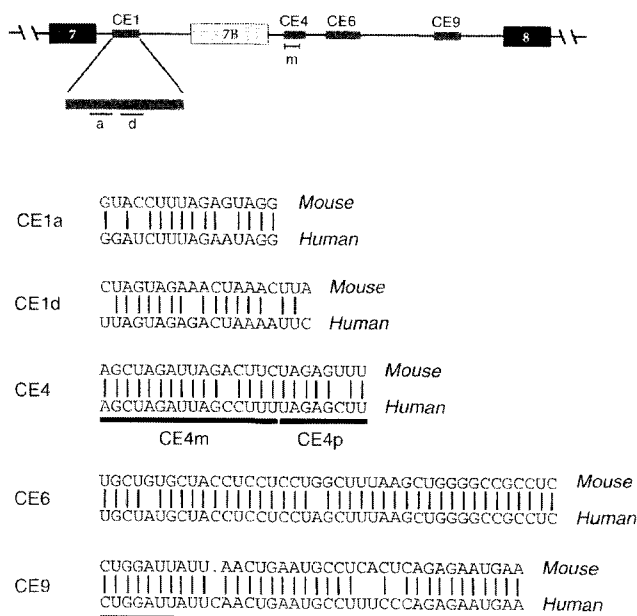


Figure 17. Éléments introniques conservés de la protéine hnRNP A1. Les boîtes représentent les exons, les lignes représentent les introns. La séquence CE1 contient les éléments CE1a et CE1d. La séquence CE6 forme une structure secondaire avec le site d'épissage 5' de l'exon 7B. L'élément CE4 comporte des sites de haute affinité pour la liaison de hnRNP A1. Tiré de Chabot et al. (2003).

CE1 à CE10 (CE, conserved elements), qui sont conservées à plus de 90%. La caractérisation de certains de ces différents éléments a permis de mieux comprendre la modulation de l'épissage alternatif par la protéine A1 sur son propre ARN messenger.

L'élément CE1 (150 nt) est

situé en amont de l'exon 7B et peut promouvoir l'exclusion de l'exon alternatif *in vivo* (Chabot et al. 1997). *In vitro*, lorsque cet élément est placé entre deux sites d'épissage

5' en compétition, il mène à une utilisation maximale (90%) du site d'épissage 5' distal (Chabot et al. 1997). Il a été montré que la protéine A1 elle-même peut se lier à certaines séquences contenues dans l'élément CE1 (CE1a et CE1d), et que la modulation de l'épissage est due à la protéine A1 (Chabot et al. 1997; Hutchison et al. 2002). D'autres sites de haute affinité pour A1 ont été identifiés en aval de l'exon 7B, dont l'élément CE4 (Chabot et al. 1997; Blanchette and Chabot 1999), et son activité dans l'épissage est similaire à l'activité de CE1 (Hutchison et al. 2002).

Les résultats obtenus ont permis d'élaborer un modèle (figure 18) expliquant l'activité de hnRNP A1 dans la modulation de la sélection des sites d'épissage. Dans ce

modèle, la liaison des molécules A1 aux éléments CE1 et CE4 contenus dans les introns flanquant l'exon alternatif serait suivie d'interactions entre ces protéines liées à l'ARN pré-messager. Cette interaction entre les molécules A1 permettrait le « looping out » l'exon alternatif 7B et par conséquent, du site d'épissage 5' proximal, menant à une répression de son usage. Ce changement de conformation de l'ARN permettrait le rapprochement spatial de sites d'épissage autrefois distants (site d'épissage 5' de l'exon 7 et site d'épissage 3' de l'exon 8), pour augmenter leur utilisation lors de l'épissage.

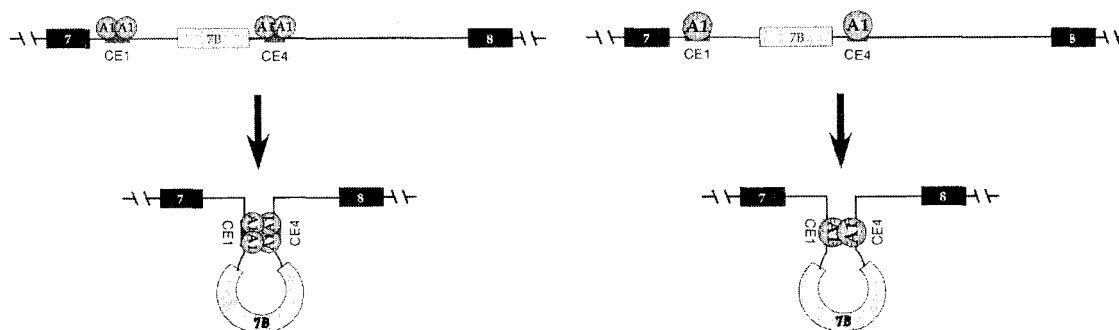


Figure 18. Modèle d'action de hnRNP A1. Une interaction entre des molécules hnRNP A1 liées aux éléments CE1 et CE4 permettrait l'exclusion de l'exon alternatif 7B par un mécanisme de « looping out ». La présence de multiples sites de liaison pour A1 pourrait permettre de lier plusieurs protéines sur un élément (gauche) ou une protéine pourrait lier deux sites adjacents, par ses deux domaines RRM (droite). Tiré de Chabot et al. (2003).

Étant donné la distribution intronique des sites de liaison pour hnRNP A1, un modèle similaire a été proposé pour l'action de A1 dans l'épissage constitutif. De plus, il a été montré que ce mécanisme de « looping out » est semblable à celui imposé par l'insertion de séquences duplexes stables dans l'ARN, et pourrait aussi agir lors de l'épissage de longs introns (Nasim et al. 2002). Le modèle proposé est en accord avec la structure du cristal de la portion UP1 de la protéine avec un oligo télomérique modèle (Ding et al. 1999). Dans ce co-cristal, UP1 existe comme un dimère et chacune des deux molécules interagit avec un site de liaison sur les oligonucléotides (Ding et al. 1999).

Pour cette raison, il est proposé que les protéines A1 liées à l'ARN par le RRM1 (Dallaire et al. 2000) pourraient par la suite interagir entre elles par leur domaine GRD, et ce complexe pourrait être stabilisé par des interactions entre les domaines RRM2 et l'ARN.

Les protéines hnRNP A1 et A2 possèdent une structure semblable et partagent 68% d'identité. Les membres de cette famille de protéine ont une activité similaire en ce qui a trait à la modulation de la sélection des sites d'épissage (Mayeda et al. 1994). Des essais de chromatographie d'affinité avec des molécules d'ADN ou d'ARN comportant des séquences de liaison de haute affinité pour hnRNP A1, permettent d'enlever de l'extrait nucléaire les protéines hnRNP A1/A1^B/A2/B1. Ce résultat suggère que chacune des protéines de cette famille est capable de lier les mêmes séquences (Caputi et al. 1999; Hutchison et al. 2002). En effet, les formes recombinantes de A1, A1^B et A2 sont toutes capables de lier le site de liaison de haute affinité de hnRNP A1 (Mayeda et al. 1994; Hutchison et al. 2002). De plus, il a été montré que chacune de ces protéines peut permettre la modulation de la sélection des sites d'épissage sur un transcrit modèle contenant des sites de haute affinité pour A1 (Hutchison et al. 2002). Cependant, certains autres sites permettant la liaison de ces protéines ne peuvent pas permettre la modulation de l'épissage, même lorsque mis dans un contexte d'épissage alternatif (Bériault et al. 2004).

En plus de son action dans l'épissage, la protéine hnRNP A1 a une fonction aux télomères. En effet, la séquence de liaison de haute affinité de hnRNP A1 (UAGGG^A/U) (Burd et Dreyfuss 1994) est la même que celle qui est retrouvée aux télomères humains (TTAGGG) (Moyzis et al. 1988; Moyzis 1991). En plus de la protéine hnRNP A1, les

autres membres de cette famille peuvent aussi lier les séquences télomériques humaines (McKay et Cooke 1992b; McKay et Cooke 1992a). La protéine hnRNP A1 peut aussi participer à la biogénèse des télomères (LaBranche et al. 1998). Dans une lignée de cellules de souris déficientes pour la production de hnRNP A1 (Ben-David et al. 1992), la taille des télomères est plus courte que dans la lignée cellulaire parentale (LaBranche et al. 1998). Le fait de restaurer la production de la protéine A1, ou sa forme amino-terminale UP1, permet le rallongement de la taille des télomères (LaBranche et al. 1998). La fraction UP1 est capable de se lier de façon préférentielle à des oligos possédant des extensions 3' comportant des séquences télomériques humaines (Dallaire et al. 2000) et le domaine RRM1 semble suffisant pour cette liaison spécifique (Dallaire et al. 2000). La liaison des protéines hnRNP A1/UP1 aux extrémités des oligos permet la protection contre des processus de dégradation (Dallaire et al. 2000). De plus, il a été montré que les protéines hnRNP A1/UP1 peuvent lier directement la composante ARN de la télomérase humaine (hTR) (Fiset et Chabot 2001), et que cette liaison est spécifique au RRM2 (Fiset et Chabot 2001), ce qui peut suggérer un rôle pour A1 dans le recrutement de la télomérase aux extrémités des chromosomes. Également, il a été observé que les niveaux de production des protéines hnRNP A1/A2 sont élevés dans plusieurs lignées de cancers humains (Patry et al. 2003; Biamonti et al. 1993; Zhang et al. 1997; Iervolino et al. 2002; Zhou et al. 1996; Sueoka et al. 1999; Satoh et al. 2000; Yan-Sanders et al. 2002; Snead et al. 2003). La réduction des niveaux cellulaires de hnRNP A1/A2 par interférence médiée par l'ARN (RNAi) provoque la mort par apoptose dans des lignées de cellules dérivées de cancers, mais pas dans des lignées de cellules normales (Patry et al. 2003). Les mêmes effets sont observés pour des lignées cellulaires de souris (Patry et al. 2004). Ces

différents résultats montrent que la protéine hnRNP A1 est importante et impliquée dans deux processus cellulaires complètement différents, dont les mécanismes moléculaires sont encore à l'étude.

Objectifs du projet

Il a été montré que la protéine hnRNP A1/UP1 a un effet sur la taille des télomères (LaBranche et al. 1998). Par contre, au début du projet, aucun indice sur le mécanisme moléculaire ou les partenaires cellulaires impliqués dans ce processus n'était connu. Au départ, le projet visait à trouver les partenaires cellulaires de hnRNP A1/UP1, permettant de mieux caractériser son action aux télomères. L'hypothèse de départ était que certains partenaires cellulaires de hnRNP A1 pourraient moduler les fonctions de la protéine, autant au niveau de son action aux télomères que dans l'épissage alternatif. En trouvant les partenaires cellulaires de hnRNP A1, il serait ainsi possible de mieux comprendre les implications de la protéine dans les deux mécanismes. Avec une approche de criblage double-hybride chez la levure, plusieurs partenaires cellulaires ont été identifiés. Cependant, aucune de ces protéines ne possédait une implication ou une localisation aux télomères.

Par la suite, le modèle de « looping out » (Blanchette et Chabot 1999) a été proposé pour expliquer les résultats obtenus *in vitro* et *in vivo* concernant la modulation de la sélection des sites d'épissage par la protéine hnRNP A1. Cependant, les mécanismes moléculaires menant à cette sélection des sites d'épissage étaient encore mal compris. Comme des résultats positifs pour une interaction entre des molécules hnRNP A1 avaient été obtenus lors du premier criblage double-hybride effectué, ce système a été utilisé pour tester certains éléments du modèle proposé. Suite à la production de protéines hnRNP A1 mutantes, le projet consistait à caractériser plus en détails les différents domaines de hnRNP A1 requis pour les étapes distinctes de la modulation de l'épissage,

d'approfondir les mécanismes moléculaires menant à cet événement, et en bout de ligne, de voir s'il est possible avec cette approche de valider le modèle proposé.

Finalement, étant donné que nous avons plusieurs transcrits pour l'étude de la modulation de la sélection des sites d'épissage, un autre projet visait à caractériser les mécanismes d'action de la protéine hnRNP A2 (Hutchison et al. 2002; Bériault et al. 2004).

CHAPITRE 1

Partenaires cellulaires de hnRNP A1/UP1 et implications dans différents processus cellulaires.

Préambule

Afin de trouver des protéines pouvant être des partenaires de hnRNP A1/UP1 dans la cellule, j'ai effectué un criblage double-hybride dans la levure. Le but était de repêcher des protéines pouvant interagir avec hnRNP A1/UP1, et ainsi obtenir plus d'informations sur son action au niveau de la biogénèse des télomères.

Par la suite, avec ce système de détection des interactions mis en place, j'ai entrepris un deuxième criblage. Celui-ci avait pour objectifs de mieux caractériser les différents domaines de la protéine hnRNP A1 impliqués dans sa dimérisation et ultimement, de valider le modèle proposé pour l'action de cette protéine dans l'épissage alternatif (Blanchette et Chabot 1999). Le criblage de perte d'interactions qui a été effectué a permis de récupérer des protéines hnRNP A1 contenant différentes mutations affectant leur capacité d'interaction entre elles. Ces protéines mutantes ont par la suite été analysées pour leur capacité à lier l'ARN. Elles ont aussi été étudiées pour leur capacité à moduler l'épissage alternatif d'un transcrit modèle, dans des extraits dépourvus de la protéine hnRNP A1 endogène.

De plus, l'utilisation du système de double-hybride et le projet sur l'interaction entre les protéines hnRNP A1 m'ont permis de collaborer avec plusieurs autres étudiants du laboratoire, travaillant sur différents aspects du projet. J'ai réalisé toutes les expériences décrites dans cette section; aucun manuscrit en préparation avec ces résultats.

MATÉRIEL ET MÉTHODES

Généralités sur le double-hybride

La méthode double-hybride chez la levure est utilisée fréquemment afin de trouver des partenaires cellulaires à une protéine d'intérêt. Brièvement, la méthode peut être récapitulée comme il est montré dans la figure 19.

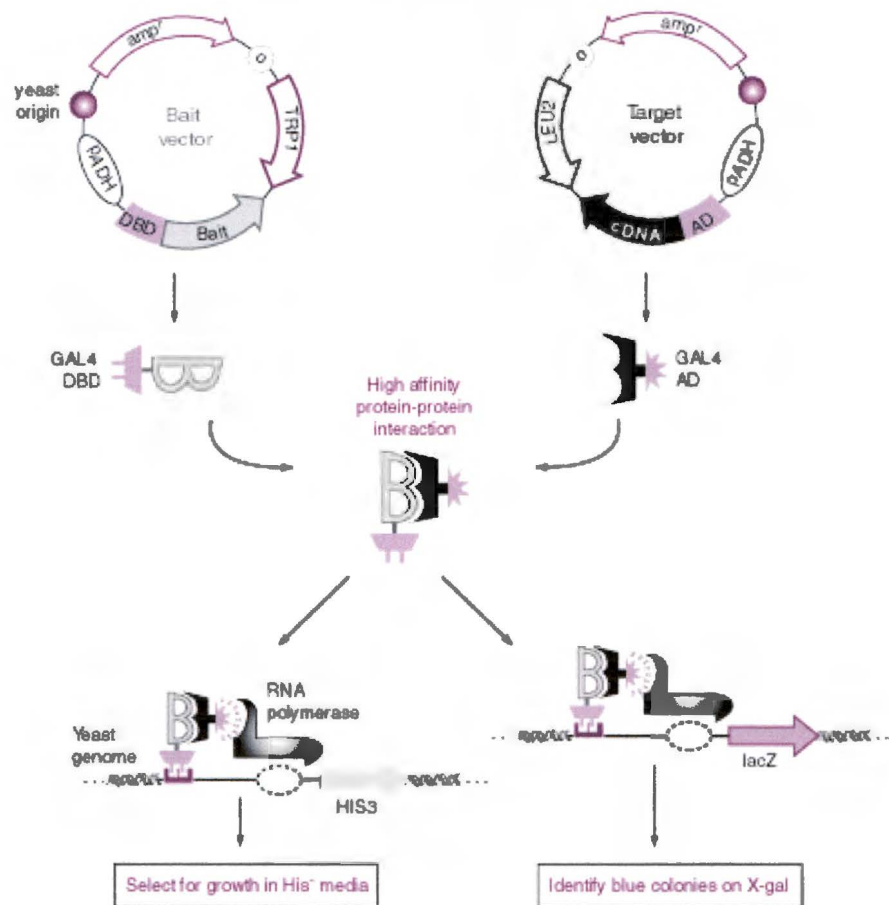


Figure 19. Double-hybride chez la levure et plasmides utilisés. Le plasmide nommé « Bait vector » est pAS2-UP1 ou pAS2-A1 : il permet la production d'une protéine de fusion GAL4 DBD-UP1 ou GAL4 DBD-A1. Le plasmide nommé « Target vector » est pGAD-GH : il contient les fragments de l'ADNc de la librairie, et permet la production de protéines de fusion GAL4 AD-protéine de librairie. S'il y a interaction entre les protéines UP1/hnRNP A1 et une protéine de la librairie, le DBD de GAL4 se lie à un site spécifique en amont d'un gène rapporteur. L'AD de GAL4 qui est mis à proximité du GAL4 DBD par l'interaction entre les protéines permet de recruter l'ARN polymérase à cet endroit, et il y a transcription du gène rapporteur inséré dans l'ADN génomique de la souche de levure. Les protéines produites à partir du gène rapporteur (marqueur nutritionnel) permettent l'identification des interactions protéine : protéine grâce à une croissance des levures sur des milieux restrictifs. Adapté de http://www.wiley.com/legacy/products/subject/life/genetics/miesfeld_figures.html.

Plasmides, souches de levure et milieux de croissance utilisés pour le criblage double-hybride

Méthodes de clonage

Tous les clonages ont été faits selon les méthodes de clonage décrites dans (Sambrook et al. 1989). Toutes les enzymes de restriction ont été utilisées selon les recommandations des manufacturiers, sauf si mentionné autrement dans le texte.

Construction des plasmides pAS2-UP1 et pAS2-A1

Le plasmide pAS2-1 a été acheté chez Clontech (PT3023-5, numéro de catalogue K1604-B). Pour la production de protéines de fusion avec le GAL4 DBD, le site NdeI du plasmide pAS2-1 a été détruit par une digestion enzymatique et remplissage à la klenow pour produire le plasmide pAS2-1A. Par la suite, les fragments BamHI/Sall ont été isolés des plasmides pGEX-A1 ou pGEX-UP1, respectivement (LaBranche et al. 1998), et introduits dans le plasmide pAS2-1A, digéré par BamHI/Sall pour produire les plasmides pAS2-A1-8 et pAS2-UP1-38D. La production des protéines de fusion a été vérifiée par isolation des protéines à partir de cultures de levures et Western Blot avec un anticorps monoclonal α -GAL4 DBD acheté chez Clontech (voir section **Isolation des protéines de levure et Western Blot**).

Construction des plasmides pGAD-UP1 et pGAD-A1

Pour produire des protéines de fusion GAL4 AD, le site SpeI du plasmide pGAD-GH (PT3062-1, numéro de catalogue K1605-1, Clontech) a été digéré, rempli à la klenow et reliqué pour donner le plasmide pGAD-2. Ce plasmide a été digéré par BamHI/EcoRI. Les fragments BamHI/EcoRI isolés des plasmides pGEX-A1 et pGEX-UP1, respectivement, ont été introduits pour donner les plasmides pGAD-A1-8 et pGAD-UP1-5. La production des protéines de fusion avec le GAL4 AD a été vérifiée par isolation des

protéines à partir de cultures de levures et Western Blot avec un anticorps polyclonal dirigé contre le GAL4 AD (don du laboratoire du Dr. S. Abou Elela).

Construction des plasmides pYES2-UP1 et pYES2-A1

[illegible]

Construction des plasmides pGEX-A1Mxx

Les constructions des différentes protéines de fusion GST-A1Mxx ont été obtenues suite à l'isolation d'un fragment BamHI/Sall d'un plasmide pAS2-A1Mxx (voir section **Oligonucléotides, génération de mutations aléatoires par PCR, « gap repair » dans une souche de levure PJ69-4a Ku80::ura3 et criblage de la perte des interactions**). Ce fragment a été inséré dans le vecteur pGEX-KG (dérivé de pGEX-2T) digéré par les enzymes BamHI/Sall. Les plasmides obtenus ont été séquencés avec l'amorce FD-1 achetée chez Gibco-BRL (5'-TCAGACTCTCGTCGGTTGTT-3'), transformés dans les bactéries BL21 pour permettre l'induction de la production des protéines de fusion GST-A1Mxx et la purification de celles-ci (voir section **Production de protéines recombinantes**).

Amplification et purification des plasmides de la librairie

La librairie utilisée dans le criblage double-hybride a été achetée chez Clontech (Human HeLa Matchmaker cDNA Library, HL4000AA, Lot 53022a) et son titre a été évalué selon les recommandations du fabricant. Les plasmides de la librairie ont été amplifiés et purifiés selon les recommandations du fabricant (Clontech Matchmaker Gal4 Two-Hybrid User manual PT3061-1).

Souche de levure utilisée pour le criblage, PJ69-4a

La souche de levure utilisée pour le criblage double-hybride est la souche PJ69-4a (don du Dr. Philip James, Université du Wisconsin) (James et al. 1996).

Milieux de croissance

Les milieux de croissance utilisés pour les levures sont des milieux de croissance standards, décrits dans (Rose et al. 1990). Les milieux de croissance utilisés pour les bactéries sont des milieux de croissance standards, décrits dans (Sambrook et al. 1989).

Protocole pour une transformation de levure à haute efficacité

Le protocole utilisé pour une transformation double-hybride à haute efficacité est celui décrit dans (Gietz et al. 1995). (Plusieurs améliorations subséquentes de ce protocole sont disponibles à l'adresse suivante; www.umanitoba.ca/faculties/medicine/biochem/gietz/Trafo.html.) Les levures transformées ont été étalées sur des boîtes Yc-Trp-Leu-His + 5mM 3-AT (3-aminotriazole) et incubées à 30°C de 5 à 7 jours. Un aliquot de la transformation a été étalé sur des boîtes Yc-Trp-Leu mises à 30°C pour 2-3 jours afin de calculer le nombre total de transformants obtenus pour chaque transformation.

Vérification des interactions entre les protéines par activation des gènes rapporteurs HIS3, ADE2, LacZ

Les boîtes Yc-Trp-Leu-His + 5mM 3-AT ont été incubées à 30°C pour permettre la croissance des levures ayant subi le protocole de transformation. Les levures récupérées suite à cette étape ont été striées sur des boîtes Yc-Trp-Leu-His-Ade et incubées à 30°C de 5 à 7 jours. Les levures qui poussent sur ces boîtes ont été striées une 2^e fois sur le même milieu de sélection (Yc-Trp-Leu-His-Ade). Les levures qui poussent sur ces boîtes après la 2^e étape de stries ont par la suite subi un essai β -Gal sur boîte (voir section **Oligonucléotides, génération de mutations aléatoires par PCR, « gap repair »**

dans une souche de levure **PJ69-4a Ku80::ura3** et criblage de la perte des **interactions**). Par la suite, les plasmides de la librairie ont été isolés et introduits dans des bactéries par la méthode de transformation par choc thermique (voir section **Isolation de l'ADN du plasmide de la librairie et transformation dans les bactéries**). Les plasmides ont été isolés des bactéries, réintroduits dans des levures PJ69-4a, PJ69-4a+pAS2, PJ69-4a+pAS2-UP1, PJ69-4a+pAS2-A1 et ces différentes souches de levures ont été étalées sur des milieux de croissance appropriés. Les colonies récupérées suite à ces transformations ont été striées sur des boîtes Yc-Trp-Leu-His-Ade et incubées à 30°C de 5 à 7 jours. Les colonies récupérées suite à cette incubation ont été soumises à un 2^e essai β -Gal sur boîtes. Certaines colonies positives ont aussi été soumises à un essai β -Gal en liquide tel que décrit dans <http://www.fhcrc.org/labs/gottschling/yeast/Bgal.html>. Le plasmide de la librairie a été récupéré à partir des colonies qui poussent sur le milieu Yc-Trp-Leu-His-Ade et qui s'avéraient positives pour l'expression de la β -Galactosidase. Les plasmides ont par la suite été séquencés (voir section **Réaction de séquençage sur l'ADN plasmidique de la librairie isolé de bactéries**) et les différentes protéines encodées ont été identifiées par utilisation du programme BLAST.

Isolation de l'ADN du plasmide de la librairie et transformation dans les bactéries

Après le criblage de la librairie, les clones positifs récupérés contenaient deux plasmides (plasmide de la librairie et plasmide pAS2-UP1) (figure 19). Pour récupérer et analyser le plasmide de la librairie, les levures positives ont d'abord été traitées pour perdre le plasmide pAS2-UP1 contenant le marqueur de sélection TRP1. Brièvement, les levures PJ69 contenant le plasmide pAS2-UP1 (TRP1) et le plasmide de la librairie

(pGAD, LEU2) ont été inoculées dans 3 ml de YEPD liquide et mises à 30°C pour 12 heures. La culture a été diluée pour obtenir un aliquot contenant ~ 300 cellules et celles-ci ont été étendues sur des boîtes YEPD. Après 3 jours de croissance à 30°C, ~ 50 cellules ont été mises en plaque sur des boîtes YEPD et réincubées à 30°C pour 2 jours. Ces boîtes ont été répliquées sur des boîtes Yc-Leu-Trp, Yc-Leu, Yc-Trp et YEPD. Après 3 jours d'incubation à 30°C, les plaques sur les différentes boîtes ont été analysées, et les clones poussant uniquement sur les boîtes Yc-Leu et YEPD ont été gardés (perte du plasmide pAS2-UP1 et présence du plasmide de la librairie). L'ADN des clones a été isolé tel que décrit dans Huberman et al. (1987), traité à la RNase A, puis purifié sur une colonne de Séphadex G50. L'ADN plasmidique a été analysé sur gel d'agarose 0,6% dans un tampon TAE et un aliquot a été transformé dans une souche de bactéries chimiocompétentes par protocole standard de choc thermique, décrit dans Sambrook et al. (1989) afin de procéder à son analyse.

Réaction de séquençage sur l'ADN plasmidique de la librairie

L'ADN a été isolé des bactéries par protocole standard de lyse cellulaire tel que décrit dans (Sambrook et al. 1989). L'ADN ainsi récupéré a été dosé et soumis à une réaction normale de séquençage avec la trousse de séquençage par l'ADN polymérase T7 (Amersham Biosciences), selon les recommandations du manufacturier. L'oligo pAS2-1 utilisé pour le séquençage du plasmide de la librairie comporte la séquence 5'-TGCAGGAATTCGGCACGAG-3' et a été commandé chez Gibco-BRL.

Oligonucléotides, génération de mutations aléatoires par PCR, « gap repair » dans une souche de levure PJ69-4a Ku80::ura3 et criblage de la perte des interactions

La souche PJ69-4a Ku80::ura3 est dérivée de la souche PJ69-4a (James et al. 1996), dans laquelle le gène *KU80* a été remplacé par le marqueur de sélection Ura3 par la méthode de délétion médiée par le PCR (Baudin et al. 1993). Cette souche ainsi produite a été transformée avec le plasmide pGAD-A1-8, avec des techniques et des milieux habituels (Rose et al. 1990).

Les oligos pAS2-FWD-1 (5'-TCATCGGAAGAGAGTAGTAACAAAG-3') et pAS2-REV (5'-ACCTACAGGAAAGAGTTACTCAAG-3') ont été commandés chez Life Technologies. Les réactions de mutagenèse par PCR avec ces oligos ont été effectuées sur 2 ng du plasmide pAS2-A1-1, avec 20 pmol de l'amorce pAS2-FWD-1 et 20 pmol de l'amorce pAS2-REV, 10 mM Tris-HCl pH 9.0, 50 mM KCl, 7 mM MgCl₂, 0,2 mM dGTP, 0,2 mM dATP, 1 mM dCTP, 1 mM dTTP et ~ 1.5 unités de Taq ADN Polymerase. Les réactions ont été mises dans le PCR et le programme suivant a été accompli : 95°C pour 3 minutes, suivi de 30 cycles de 95°C pour 1 minute, 45°C pour 1 minute et 72°C pour 2 minutes, puis une étape d'élongation finale à 72°C pour 15 minutes. Le produit de PCR obtenu (pAS2-A1 avec des mutations aléatoires) a été purifié sur une colonne Amicon ® Microcon®-PCR de la compagnie Millipore, selon les recommandations du manufacturier.

Par la suite, environ 1 µg du fragment purifié de pAS2-A1 contenant les mutations aléatoires et environ 900 ng du vecteur pAS2-1A linéarisé par BamHI ont été co-transformés dans la souche PJ69-4a Ku80::ura3 contenant le plasmide pGAD-A1-8, selon un protocole de transformation de haute efficacité décrit dans Gietz et al. (1992).

Les levures transformées ont été mises sur des boîtes Yc-Trp-Leu, pour sélectionner la présence du plasmide pGAD-A1-8 (LEU2) et des événements de recombinaison homologue entre les séquences du plasmide linéaire pAS2-1A (TRP1) et les fragments amplifiés à partir du plasmide pAS2-A1, reformant des plasmides circulaires (figure 23).

Les levures récupérées suite à cette sélection ont été déposées sur des boîtes Yc-Leu-Trp et ont été soumises à un essai β -Gal sur boîte, adapté du protocole de Clontech dans Yeast Protocols Handbook PT3024-1. Brièvement, une solution de 10% (w/v) X-Gal dans du diméthyl formamide est ajoutée à 10 ml d'une solution 1% agarose, 0.2 M NaH_2PO_4 , 0.3 M Na_2HPO_4 et 0.0005% (w/v) SDS. Ce mélange est uniformément réparti sur les boîtes où les levures ont poussé puis ces boîtes sont incubées à 37°C pour environ 1 heure, et à 4°C pour 10 minutes. Suite à ces incubations, un aperçu relatif de l'intensité des interactions entre les protéines de fusion est obtenu par comparaison de couleur (bleu foncé, bleu pâle, blanc). Les colonies restées blanches après ce traitement ont été gardées (perte de l'interaction entre les protéines GAL4 DBD-A1 contenant les mutations aléatoires et GAL4 AD-A1), mises en culture dans un milieu liquide Tc-Leu-Trp, puis soumises au protocole d'extraction des protéines (voir section **Isolation des protéines à partir de cultures de levures et procédure de Western Blot**) et à une analyse par western blot avec l'anticorps α -GAL4 DNA-BD, afin de ne garder que les plasmides pAS2-A1-Mxx contenant un fragment mutagénisé menant à la production d'une protéine de fusion complète GAL4 DBD-A1.

Les plasmides récupérés ont été soumis aux autres étapes de l'analyse.

Isolation des protéines à partir de cultures liquides de levures et procédure de Western Blot

Le protocole SDS-Urée utilisé pour préparer des extraits de protéines de levure est celui décrit dans « Clontech Yeast Protocols Handbook » PT3024-1, version PR67090. Des aliquots pris à partir des extraits de protéines finaux ont été chargés sur un gel de protéines de 12,5% (polyacrylamide/bisacrylamide 29 :1) tel que décrit dans (Sambrook et al. 1989).

L'anticorps monoclonal α -GAL4 DBD a été acheté chez Clontech (5399-1) et utilisé comme anticorps primaire dilué à 1:500 dans du TBST+ 5% (w/v) lait. L'anticorps secondaire α -mouse IgG utilisé a été acheté chez SantaCruz Biotechnology et dilué à 1:1000 dans du TBST+ 5% (w/v) lait. Une réaction de chemiluminescence (ECL) typique a été faite et suivie d'une exposition de 5 minutes sur un film Kodak XAR.

Isolation des protéines à partir de cultures liquides de levures et gel de retardement (« gel shift assay »)

Un protocole adapté de Berman et al. (1986) a été utilisé pour la production des extraits de levures. Brièvement, les cultures de levures sont centrifugées et lavées avec de l'eau. Le culot est resuspendu dans 650 μ l de tampon A (50 mM KCl, 50 mM Hepes pH 7.5, 5 mM EDTA, 5% (v/v) glycérol) et 100 μ l de la solution d'inhibiteur de protéases sont rajoutés (completeTM EDTA-free protease inhibitor cocktail, Boeringher Mannheim). Les culots sont transférés dans des tubes en verre, gardés sur la glace. Un volume de billes de verre (0.45-0.55 mm, Sigma-Aldrich) est rajouté au mélange sur la glace. Les tubes sont vortexés 1 minute à 4°C, puis remis sur la glace pour 1 minute, 5 fois. Les

extraits sont récupérés et centrifugés 20 minutes à 4°C, et le surnageant est utilisé dans les essais de formation de complexe puis conservé à -80°C.

L'oligo TS10 (5'-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG-
TTAGGGTTAGGGTTAGGGTTAGGG-3') utilisé a été obtenu de Gibco BRL. Les
réactions de marquage de l'oligo en 5' ont été effectuées comme dans Sambrook et al.
(1989).

Les réactions de formation de complexe et de gel de retardement ont été faites comme décrit dans Dallaire et al. (2000).

Marquage d'oligo ARN et gel de retardement avec des protéines recombinantes (« gel shift assay »)

Les réactions de marquage d'un oligo ARN, de liaison de l'ARN avec des protéines recombinantes et de gel de retardement ont été effectuées comme décrites dans (Hutchison et al. 2002), avec les oligos ARN CE1a (5'-UACCUUUAGAGUAGGC-3') ou CE4 (5'-AGCUAGAUAUAGACUUCUAGAGUUU-3') achetés chez Dharmacon Research, Inc.

**Extraits nucléaires, extraits nucléaires dépourvus des protéines hnRNP A/B,
transcrits modèles et réactions d'épissage *in vitro***

Les constructions des transcrits modèles utilisés, soit les transcrits 86.6 (pC5' 4/4) et 68.1RV (pC5' -/-) ont été décrits dans Blanchette et Chabot (1999). Les substrats ont été produits en linéarisant les plasmides par digestion enzymatique avec ScaI, et ont été

transcrits avec l'ARN polymérase T3 (Amersham Biosciences) en présence de ribonucléotides, d'un analogue de cap, et de [α -³²P]UTP (Amersham Biosciences).

La purification des transcrits d'ARN a été effectuée sur un gel 5% polyacrylamide dénaturant.

Les extraits nucléaires de cellules HeLa, CB3C7 et CB3C7-20 ont été préparés selon le protocole décrit dans Dignam et al. (1983).

Une colonne d'agarose hydrazide acide adipique (Amersham Biosciences) sur laquelle a été couplé un ARN contenant des sites de liaison de haute affinité pour les membres de la famille hnRNP A/B (CE1a ou CE4) a été utilisée pour produire les extraits nucléaires dépourvus de ces protéines, comme décrit dans Hutchison et al. (2002). Ces extraits nucléaires dépourvus des protéines de la famille hnRNP A/B et les protéines ainsi fixées sur la colonne d'agarose ont été analysés par Western Blot comme décrit dans Hutchison et al. (2002).

Les extraits nucléaires ont été utilisés par la suite dans les réactions d'épissage, selon le protocole décrit dans Chabot et al. (1997).

Production de protéines recombinantes

Les protéines de fusion GST-A1, GST-A2 et GST-A1Mxx ont été produites en suivant la procédure II décrite dans Fiset et Chabot (2001). La concentration des protéines recombinantes a été mesurée par Bradford (Bio-Rad) en utilisant des dilutions en série de BSA comme standard de mesure. Ces protéines recombinantes ont par la suite été rajoutées à un extrait nucléaire dépourvu des protéines de la famille hnRNP A/B, et

des réactions d'épissage *in vitro* ont été effectuées, comme décrit dans Hutchison et al. (2002).

Souche de levure ou lignée cellulaire	Caractéristiques des souches de levure ou des lignées cellulaires	Provenance
MW35a	MATa , <i>ade2</i> , <i>ade5</i> , <i>can1</i> ^R , <i>cyh2</i> ^R , <i>lys5</i> , <i>ura3-52</i> , <i>trp1Δ</i> , <i>leu2-3</i> , 112, <i>his3-200</i> , <i>gal++</i> Souche de levure permettant la surexpression de protéines (induction de leur production par incubation dans un milieu de croissance contenant du galactose)	Dr. R.J. Wellinger
PJ69-4a	MATa <i>trp1-901 leu2-3, 112 ura3-52 his3-200 gal4Δ gal80Δ LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacZ</i> Souche de levure utilisée pour le criblage double-hybride. Elle contient 3 gènes rapporteurs, chacun sous le contrôle d'un promoteur inducible.	Dr. P. James (James et al. 1996)
PJ69-4a, Ku80::ura3	Même souche que celle décrite plus haut. Elle contient aussi la délétion du gène <i>Ku80</i> , afin de minimiser les événements de NHEJ lors de son utilisation pour le « gap repair ».	Dr. P. James (James et al. 1996)
HeLa	Cellules épithéliales dérivées de la lignée parentale de HeLa, adénocarcinome, expriment HPV 18, bas niveau de p53, niveau normal de Rb	Dr. B. Chabot
CB3C7	Lignée cellulaire érythroleucémique de souris ayant une insertion rétrovirale menant à une inactivation de l'allèle <i>Hnrpa1</i> alors que l'autre allèle a été perdue. Ces cellules expriment environ 200-500 fois moins de transcrit <i>Hnrpa1</i> et pas de protéine hnRNP A1 détectable.	Dr. B. Chabot (Ben-David et al. 1992)
CB3C7-20	Lignée cellulaire dérivée de CB3C7, dans laquelle l'expression de hnRNP A1 est rétablie. Ces cellules ont donc un niveau normal de protéine A1.	Dr. B. Chabot (Yang et al. 1994).

Tableau 3. Caractéristiques des souches de levure et lignées cellulaires utilisées.

RÉSULTATS

Criblage double-hybride avec la protéine hnRNP A1/UP1 utilisée comme appât

La protéine hnRNP A1 comporte 320 acides aminés (Buvoli et al. 1988) et est organisée en deux domaines spécifiques. Le domaine N-terminal (nommé UP1) comporte 196 acides aminés et il contient deux domaines RRM (Dreyfuss et al. 1993). La portion C-terminale de la protéine contient un domaine riche en glycine (GRD) et semble avoir un rôle dans l'interaction avec d'autres protéines (Casas-Finet et al. 1993; Shamoo et al. 1994; Cartegni et al. 1996).

L'intérêt de trouver des partenaires cellulaires de hnRNP A1/UP1 vient de son action dans la biogénèse des télomères (LaBranche et al. 1998). En effet, ceci constitue un nouveau rôle pour cette protéine. En trouvant des partenaires d'interaction de hnRNP A1/UP1, il serait peut-être possible de mieux comprendre l'implication télomérique de cette protéine. Les résultats publiés à ce sujet montrent qu'une forte diminution de l'expression de hnRNP A1/UP1 dans une lignée cellulaire est associée à un raccourcissement de la longueur des télomères (LaBranche et al. 1998). De plus, dans ces mêmes cellules déficientes pour l'expression de hnRNP A1 (tableau 3), l'introduction exogène de la portion N-terminale de hnRNP A1 (UP1) permet un rallongement des télomères (LaBranche et al. 1998). Deuxièmement, les protéines hnRNP A1 et UP1 peuvent chacune lier directement et spécifiquement des oligonucléotides comportant des séquences télomériques humaines *in vitro* (LaBranche et al. 1998). Cependant, seulement la portion UP1 semble capable de récupérer l'activité de l'enzyme télomérase détectée par essai TRAP à partir d'un lysat cellulaire (LaBranche et al. 1998). Ces données suggèrent donc un rôle de la protéine hnRNP A1/UP1 dans un processus relié aux

télomères, en plus de son rôle connu et documenté dans la modulation de l'épissage alternatif. Dans le but de mieux comprendre ses différents modes d'action, le projet initial était de découvrir les partenaires cellulaires possibles de hnRNP A1/UP1.

Dans cette optique, nous avons utilisé la partie UP1 de hnRNP A1 dans un criblage double-hybride dans la levure. Même s'il est peu probable que la protéine UP1 existe dans les cellules de mammifères *in vivo*, ce domaine garde l'habileté de la protéine complète à promouvoir l'élongation des télomères et peut interagir avec la télomérase (LaBranche et al. 1998). Elle demeure donc un bon modèle afin de développer des systèmes pour l'étude de la modulation de la longueur des télomères.

Le système double-hybride (Fields et Song 1989) dans la levure est largement utilisé comme une première approche pour découvrir les partenaires cellulaires d'une protéine d'intérêt. La protéine UP1 a été fusionnée au domaine de liaison à l'ADN de la protéine GAL4 menant à la production du plasmide pAS2-UP1, qui fut introduit dans la souche de levure PJ69-4a (James et al. 1996). Cette souche possède 3 gènes rapporteurs distincts (HIS3, ADE2, LacZ), sous le contrôle de promoteurs différents dont la transcription est induite par la protéine GAL4. Ces gènes rapporteurs permettent de visualiser s'il y a interaction entre les différentes protéines. En effet, pour permettre la transcription des gènes rapporteurs, il faut que les deux domaines distincts, mais physiquement séparés de la protéine GAL 4 (c'est-à-dire le domaine de liaison à l'ADN (DBD), et le domaine d'activation de la transcription (AD)) soient mis à proximité l'un de l'autre grâce à l'interaction entre deux protéines de fusion (figure 19).

Les contrôles usuels pour un criblage ont été complétés avec la protéine de fusion GAL4 DBD-UP1 (e.g. toxicité de la protéine de fusion produite, quantité de protéine de

fusion produite, production d'une protéine de fusion complète, absence d'activation des gènes rapporteurs par la protéine de fusion elle-même; données non-montrées). Par la suite, nous avons vérifié si la protéine de fusion GAL4 DBD-UP1 produite dans la levure garde sa capacité de lier un oligonucléotide contenant des répétitions télomériques humaines (figure 20). Dans cet essai, des extraits de protéines de levures sont produits et incubés avec un oligo télomérique radiomarké, en présence d'inhibiteur de protéases et de compétiteur. La formation de complexes entre les protéines de fusion produites dans la levure et l'oligo radiomarké est permise 30 minutes à température ambiante. Les différents complexes sont par la suite séparés sur un gel de polyacrylamide non-dénaturant. Dans la figure 20, plusieurs complexes protéiques non-spécifiques sont présents dans toutes les pistes où les extraits de levures ont été incubés avec l'oligo TS10 (flèches blanches). Cependant, dans les pistes 3 à 6 de la figure 20 qui montrent des extraits protéiques contenant la protéine de fusion GAL4 DBD-UP1 (quatre clones indépendants), un complexe supplémentaire est observé (flèche noire). Ce complexe est visible seulement dans les pistes où la protéine GAL4 DBD-UP1 est exprimée. Ce complexe ne migre cependant pas à la même hauteur que celui observé entre la protéine UP1 clivée et l'oligo TS10 (piste 8). Cette différence de migration peut être expliquée par une multimérisation possible des protéines UP1 clivées sur l'oligo TS10, donnant lieu à un plus gros complexe qui est plus retardé dans le gel. Ainsi, la protéine GAL4 DBD-UP1 est produite dans les levures, et la portion GAL4-DBD en 5' de la protéine UP1 n'entrave pas l'interaction avec un oligo télomérique.

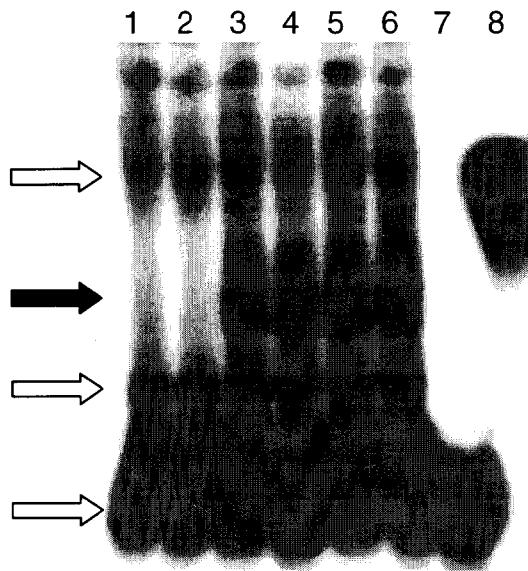


Figure 20. La protéine GAL4 DBD-UP1 peut lier un oligo télomérique humain. Gel de retardement entre les protéines GAL4 DBD-UP1 produites chez la levure et l'oligo TS10 radiomarqué (10 répétitions télomériques humaines). Les extraits protéiques ont été faits selon la méthode d'extraction adaptée de (Berman et al. 1986). La formation des complexes a été laissée à 23°C pour 30 minutes, dans un tampon contenant un mélange d'inhibiteurs de protéases et un compétiteur. Les complexes ont été séparés sur un gel polyacrylamide non-dénaturant 6%, 2 heures à 200V. La flèche noire représente un complexe spécifique entre la protéine GAL4 DBD-UP1 et l'oligo TS10. Les flèches blanches représentent des complexes non-spécifiques entre des protéines de levure et l'oligo TS10. Pistes : 1) Souche de levure PJ69 2) PJ69 + pAS2 3) PJ69 + pAS2-UP1-4D 4) PJ69 + pAS2-UP1-23D 5) PJ69 + pAS2-UP1-38D 6) PJ69 + pAS2-UP1-44D 7) oligo TS10 8) protéine recombinante GST-UP1, clivée à la thrombine.

La souche de levure PJ69-4a exprimant la protéine de fusion GAL4 DBD-UP1 a été utilisée ensuite pour le criblage d'une librairie d'ADNc provenant de cellules humaines HeLa S3. Cette librairie comporte environ 6×10^6 clones indépendants, et environ 93% des plasmides de cette librairie possèdent un insert d'ADN, dont la taille moyenne est de 1.5 kb.

Un total d'environ 25,5 millions de transformants a été criblé par transformation de levure à haute efficacité, ce qui correspond à environ 4 fois le nombre de clones indépendants compris dans la librairie utilisée. La sélection initiale des clones a été faite sur un milieu Yc-Trp-Leu-His + 5 mM 3-AT. Ce milieu permet d'abord de sélectionner pour les transformants ayant incorporé les deux plasmides. De plus, il permet de sélectionner les transformants qui montrent une interaction entre les protéines permettant l'activation du gène rapporteur HIS3. Le composé 3-aminotriazole agit comme un inhibiteur compétitif de la protéine His3p de levure. Il est utilisé afin d'inhiber la croissance de levures exprimant de bas niveaux de His3p de façon non-spécifique. L'expression de bas niveaux de protéine dans la levure peut être due à la présence d'un

promoteur GAL4 contenant des séquences de liaison moins spécifiques, permettant l'activation du gène rapporteur de façon peu contrôlée. L'utilisation de ce réactif est donc utile pour diminuer le taux de récupération de clones faussement positifs pour des interactions entre les protéines.

Sur les 25,5 millions de transformants criblés, 4175 se sont révélés être positifs pour l'expression de la protéine His3p. Ces colonies positives ont été striées sur un milieu Yc-Trp-Leu-His-Ade. Cette 2^e sélection permet de voir si l'interaction entre les deux protéines dans la souche de levure permet l'activation d'un 2^e gène rapporteur. L'activation de celui-ci requiert une interaction efficace entre les protéines, à cause de la spécificité de la séquence de liaison pour GAL4 contenue dans son promoteur. Ainsi, des 4175 clones montrés HIS+, seulement 780 étaient aussi ADE+. De plus, dans certains d'entre eux, les interactions étaient vraiment plus faibles que d'autres, les colonies ne poussant que dans la première strie déposée sur la boîte de culture. Pour cette raison, une deuxième série de stries sur le même milieu de sélection (Yc-Trp-Leu-His-Ade) a été faite. Des 780 clones ADE+ du départ, seulement 440 clones pouvant soutenir une croissance sur ce milieu de sélection ont été récupérés.

L'activation du 3^e gène rapporteur permet une analyse plus quantitative de l'interaction entre les protéines. En effet, il est possible de faire un test colorimétrique sur boîtes de croissance pour doser l'efficacité relative des interactions protéiques. Ainsi, sur les boîtes où les levures ont poussé on ajoute une solution de phosphate de sodium et de SDS. Ceci permet de perméabiliser les parois des cellules pour l'entrée du substrat requis pour l'essai. En plus, ce tampon augmente le pH du milieu pour permettre à la β -galactosidase formée d'être fonctionnelle. Le substrat (X-Gal) de l'enzyme est introduit

dans le milieu. Un mélange contenant de l'agarose est versé sur les boîtes afin de former une 2^e couche sur les levures. S'il y a interaction entre deux protéines dans la levure, il y a activation de la transcription du gène rapporteur encodant la β -galactosidase. Cette enzyme peut catalyser l'hydrolyse de son substrat X-Gal, ce qui produit le colorant 5-bromo-4-chloro-indole qui est détecté facilement par sa couleur bleue. L'efficacité de l'interaction entre les protéines pour l'activation du promoteur de LacZ a été évaluée. Les interactions les plus significatives ont été gardées, ce qui correspond à 181 clones positifs permettant l'activation des 3 gènes rapporteurs dans la levure.

Chacun de ces 181 clones HIS+/ADE+/LacZ+ contient un plasmide de la librairie. Afin de connaître l'identité de la protéine encodée par chaque plasmide, ceux-ci doivent être séquencés. Cependant, les interactions entre les protéines qui ont été observées doivent aussi être vérifiées une seconde fois. Pour y arriver, les différents plasmides de la librairie ont été récupérés des levures, transformés dans des bactéries, et isolés pour être réintroduits dans différentes souches de levures. Cette étape permet de vérifier la reproductibilité des interactions et la dépendance de la présence des deux plasmides (et donc des deux protéines) pour l'interaction. Les mêmes tests de croissance que ceux décrits précédemment ont été effectués sur les 181 différents clones. De ces 181 clones, 68 clones ont montré une interaction entre protéines qui était dépendante de la présence de la protéine UP1 et qui était reproductible. Ce résultat correspond à moins de 0.00027% du total des transformants criblés avec cette approche de double-hybride. Les plasmides récupérés après cette 2^e série de sélections ont tous été séquencés partiellement. Les séquences obtenues ont été comparées aux séquences du génome

humain (GenBank) en utilisant le programme BLAST du site du NCBI. Le tableau 4 résume les différents résultats obtenus avec ce criblage.

Protéine identifiée (BLAST avec Gen Bank)	Nb. de clones indépendants récupérés	Pourcentage des clones positifs totaux
RBM 6 / g-16 / DEF 3 Lung cancer Ag	24	35%
hnRNP A1	15	22%
hnRNP M4	5	7,5%
SC 35	4	6%
PUR A	4	6%
GRSF1	4	6%
IMP 3 / KOC 1	4	6%
protéines inconnues	3	4,5%
CpG island	2	3%
hnRNP D	1	1,5%
Cytokeratin 8	1	1,5%
Lipocortin	1	1,5%

Tableau 4. Synthèse des protéines pouvant interagir avec la protéine UP1. Les différentes protéines identifiées sont aussi capables d'interagir avec la protéine hnRNP A1, par essai double-hybride.

Tous les plasmides récupérés par le criblage double-hybride avec UP1 ont été transformés dans des levures PJ69-4 +pAS2-A1. Les levures ont subi les mêmes tests que ceux décrits précédemment afin de vérifier s'il y a aussi une interaction avec la protéine hnRNP A1 complète. Les transformants testés se sont tous révélés positifs pour une interaction avec la protéine hnRNP A1. De plus, l'interaction entre les protéines encodées par les plasmides pGAD-A1 et pAS2-A1 s'est avérée beaucoup plus forte que l'interaction détectée au départ entre les protéines encodées par les plasmides pGAD-A1 et pAS2-UP1. Ces conclusions ont été tirées suite à la facilité de croissance des souches sur le milieu de sélection Yc-Trp-Leu-His-Ade (formation de colonies isolées).

Globalement, les protéines récupérées par ce criblage double-hybride peuvent être catégorisées en 2 groupes. Premièrement, un groupe contenant des protéines qui peuvent

lier l'ARN. Deuxièmement, un groupe qui contient des protéines pouvant être impliquées aux télomères.

Dans 24 cas sur 68, la portion de séquence contenue dans le plasmide récupéré permettait d'identifier la protéine humaine g-16/ lung cancer antigen NY-LU-12/ RNA binding motif protein 6 (RBM6)/ RNA binding protein DEF3 (Timmer et al. 1999a; Drabkin et al. 1999; Timmer et al. 1999b). La séquence amino-terminale contenue dans ces clones donne toujours l'identification de ces 4 protéines avec la même spécificité. Peu de données sont connues sur RBM 6 et DEF 3. Par contre, g-16 et Ag NY-LU-12 sont dérivés du même gène situé sur le chromosome 3p21, qui correspond au locus d'un gène suppresseur de tumeur de poumons (Gure et al. 1998; Timmer et al. 1999b; Timmer et al. 1999a; Drabkin et al. 1999). La protéine contient 2 domaines liant l'ARN, un signal de localisation nucléaire et un motif doigt de zinc pour la liaison à l'ADN (Gure et al. 1998; Drabkin et al. 1999).

Dans 15 cas sur 68, la séquence qui est récupérée suite au séquençage de l'ADNc du plasmide récupéré donne l'identification de la protéine hnRNP A1. Des études sur les interactions de la protéine hnRNP A1 ont par ailleurs montré des résultats similaires à ceux obtenus par notre criblage. Il a été montré *in vitro* par essai de colonnes GST que des protéines de fusion GST-A1 peuvent former des complexes homodimériques et hétérodimériques avec d'autres hnRNPs (Cartegni et al. 1996). Les protéines GST-UP1 ont une affinité réduite pour les interactions avec les autres protéines testées (Cartegni et al. 1996). Des essais *in vivo* en utilisant le système double-hybride ont montré des interactions homodimériques pour hnRNP A1 et hnRNP A2 et hétérodimériques entre A1 et A2 (Cartegni et al. 1996). L'interaction détectée entre UP1 et hnRNP A1 ou hnRNP

A2 est très faible (Cartegni et al. 1996). Suite à ces résultats, les auteurs ont suggéré que le domaine riche en glycine est responsable de l'interaction observée entre les protéines hnRNP A1 (Cartegni et al. 1996). Cette interaction semble dépendante de la présence de ce domaine, et différentes portions de ce domaine pourraient interagir ensemble (Cartegni et al. 1996).

Dans 5 cas sur 68, la séquence qui est récupérée mène à l'identification de la protéine hnRNP M4. Celle-ci fait partie de la famille des hnRNPs. Elle peut lier l'ARN *in vivo* (Datar et al. 1993), semble être impliquée dans l'épissage *in vitro* (Datar et al. 1993), et peut également permettre l'inhibition de l'épissage durant des chocs thermiques (Gattoni et al. 1996). Elle contient 3 domaines de liaison à l'ARN et une région riche en glycine-méthionine dans sa portion carboxy-terminale (Gattoni et al. 1996).

Dans 4 cas sur 68, la séquence obtenue nous permet d'identifier un facteur essentiel d'épissage, SC35 (Fu et Maniatis 1990). Ce facteur fait partie de la famille des protéines SR (Tacke et Manley 1995; Tacke et Manley 1999) et possède un motif de liaison à l'ARN et un domaine carboxy-terminal riche en sérine/arginine (Fu et Maniatis 1992b). Il est requis pour la formation du complexe A lors de l'épissage des ARNs pré-messagers, plus particulièrement pour l'interaction des snRNPs U1 et U2 avec l'ARN pré-messager (Fu et Maniatis 1992a). Ce résultat est en accord avec une étude précédente qui montre que des interactions de hnRNP A1 avec certaines protéines SR sont possibles (Cartegni et al. 1996).

Ce premier ensemble forme le noyau du groupe de protéines interagissant avec UP1, dont l'interaction pourrait être dépendante de la présence d'un ARN. Cependant, la méthode de criblage utilisée ne permet pas de différencier entre une interaction directe

entre les protéines ou une interaction médiée par un autre facteur cellulaire (soit ARN ou autre protéine).

Le deuxième groupe de protéines interagissant avec hnRNP A1/UP1 est constitué de protéines ayant possiblement une action aux télomères. La protéine hnRNP A1 peut aussi être considérée comme faisant partie de ce 2^e groupe.

Dans 4 cas sur 68, la séquence récupérée du plasmide permet d'identifier la protéine PUR-A (purine rich element binding protein A), qui peut lier des séquences riches en purine, donc possiblement des séquences télomériques.

Dans 4 cas sur 68, la séquence trouvée mène à l'identification de la protéine GRSF 1 (G-rich sequence binding factor 1), qui peut aussi lier des séquences riches en guanines, donc possiblement des séquences télomériques.

Dans 1 cas sur 68, la protéine identifiée à partir de la séquence du plasmide, est hnRNP D. Il a récemment été proposé que la protéine hnRNP D puisse avoir une action possible aux télomères (Eversole et Maizels 2000), comme il a été montré aussi pour la protéine hnRNP A1. HnRNP D est capable de lier le simple-brin télomérique humain de façon spécifique (Kajita et al. 1995; Eversole et Maizels 2000), est reconnue pour avoir une activité de déstabilisation de structures où il y a pairage de G (G quartets ou G-G pairing) (Dempsey et al. 1999; Eversole et Maizels 2000) et peut avoir une interaction spécifique avec la télomérase (Eversole et Maizels 2000). Cette protéine possède 2 domaines de liaison à l'ARN, une région carboxy-terminale riche en glycine/arginine et peut aussi lier des sites 3' d'épissage in vitro (Kajita et al. 1995; Liu et Dreyfuss 1995).

Seulement 3 séquences n'ont pas donné de résultats lors de la recherche des protéines encodées dans GenBank, probablement parce que la portion du génome dans

laquelle elles sont situées n'avait pas encore été séquencée lors de la recherche. Les autres protéines qui ont été repêchées sont peu connues et leur liaison possible avec UP1 est peu claire. Certaines de ces protéines reviennent dans des tableaux de faux positifs les plus souvent récupérés lors de criblages double-hybride, et par la suite, nous ne les avons plus considérées.

Analyse d'interactions entre la protéine UP1 et des protéines connues pour leur action et leur localisation aux télomères humains

Une approche visant l'analyse d'interactions potentielles entre hnRNP A1/UP1 et des protéines télomériques connues a aussi été entreprise. Les plasmides pGAD-TRF2 (don du laboratoire du Dre. Titia de Lange) et pGAD-TP1 (don du laboratoire du Dre. Lea Harrington) ont été transformés chacun dans la souche de levure PJ69-4a+pAS2-UP1. Les colonies résultantes ont été striées sur des milieux de croissance permettant l'analyse des interactions. Aucune croissance n'a été obtenue sur des boîtes Yc-Trp-Leu-His pour ces différentes constructions, même si les protéines de fusion étaient exprimées à de hauts niveaux dans les levures (données non-montrées). La même procédure a été testée dans une souche de levure PJ69-4a + pAS2-A1, et les analyses se sont aussi avérées négatives pour une interaction possible avec ces protéines.

Analyse des différents domaines de la protéine hnRNP A1 requis pour l'interaction efficace entre les molécules hnRNP A1

Le criblage effectué auparavant a permis de mettre en évidence une forte interaction de la protéine hnRNP A1 avec elle-même, confirmant les résultats précédents

(Cartegni et al. 1996). Par contre, aucune protéine ayant une action ou une localisation précise aux télomères n'a été récupérée de ce criblage, à part hnRNP A1 elle-même. Étant donné ces résultats, nous avons concentré nos efforts sur cette protéine afin de tenter de mieux comprendre son mode d'action. En effet, hnRNP A1 possède une activité dans un autre processus biologique, soit la modulation de la sélection des sites d'épissage (Mayeda et Krainer 1992; Mayeda et al. 1993; Mayeda et al. 1994; Caceres et al. 1994; Yang et al. 1994; Chabot et al. 1997; Jiang et al. 1998; Bai et al. 1999; Blanchette et Chabot 1999; Hutchison et al. 2002; Nasim et al. 2002). Un modèle pour expliquer le mode d'action de hnRNP A1 dans ce mécanisme a été proposé (Blanchette et Chabot 1999) (figure 18), et implique une interaction entre des protéines hnRNP A1. Le modèle postule que la présence de A1 à des sites de liaison de haute affinité présents sur l'ARN pré-messager permet un rapprochement de sites d'épissage auparavant distancés, grâce à une interaction entre au moins deux protéines hnRNP A1 présentes sur l'ARN pré-messager (figure 18) (Blanchette et Chabot 1999).

Le système double-hybride utilisé auparavant a réussi à montrer que l'interaction entre au moins deux protéines A1 est détectable. Il est donc possible de tester si cette interaction joue un rôle important dans la sélection des sites d'épissage. Aussi, ce système d'analyse des interactions peut servir dans la caractérisation des différents domaines protéiques impliqués dans l'interaction. Ceci peut être possible suite à la création de protéines mutantes et de protéines comportant différentes délétions et à l'analyse des interactions de ces protéines. Nous avons décidé d'approfondir les mécanismes moléculaires par lesquels la protéine hnRNP A1 peut moduler l'épissage alternatif.

La première approche utilisée a consisté à créer des troncations de la protéine hnRNP A1 en ses différents domaines. Par la suite, les interactions entre les différentes constructions ont été analysées via le système double-hybride. L'ADNc de A1 et ses deux portions (UP1 et GRD) ont été clonés dans les vecteurs d'expression de levure pAS2-1 (Clontech) et pGAD (Clontech). Ceci permet la production des protéines de fusion avec le DBD de la protéine GAL4 et l'AD de la protéine GAL4, respectivement. Différentes combinaisons de ces protéines de fusion ont été exprimées dans la levure PJ69-4a (James et al. 1996). Les essais de croissance sur les milieux de sélection ont été faits (comme décrits auparavant) pour détecter les interactions entre les protéines. La croissance sur le milieu Yc-Trp-Leu-His est efficace pour les protéines hnRNP A1 complètes (figure 21 A, rangée du milieu et colonne de gauche) et pour les constructions contenant les portions du GRD de hnRNP A1 (figure 21A, rangée du milieu et colonne du milieu). Cette croissance semble plus difficile sur ce même milieu pour les constructions UP1/ UP1 (figure 21A, rangée du milieu et colonne de droite). Cependant, l'activation de la transcription du gène ADE2 ne permet la croissance sur des boîtes Yc-Trp-Leu-His-Ade que pour la combinaison A1/ A1 (figure 21A, rangée du bas et colonne de gauche), ce qui montre une interaction efficace entre les protéines. Nous avons aussi vérifié l'activation du 3^e gène rapporteur présent dans la souche de levure par des essais de quantification de l'activité β -Galactosidase sur des cultures liquides (figure 21B). Les valeurs obtenues montrent que l'interaction entre les protéines est la plus forte quand les protéines hnRNP A1 sont complètes. La combinaison impliquant seulement les domaines GRD montre une interaction qui semble environ 5 fois plus faible que celle entre les protéines A1, pour l'activation de LacZ. Pour ce qui est de la combinaison UP1/UP1, le niveau d'expression

de LacZ semble réduit d'environ 20 fois quand on le compare à celui de la combinaison A1/A1. Ces résultats confirment ceux obtenus auparavant (Cartegni et al. 1996) et donnent de nouveaux détails quant à l'importance des différentes portions de la protéine A1 pour l'interaction entre les protéines. Ils montrent que la portion GRD de hnRNP A1 peut interagir avec elle-même, mais moins efficacement que ce qui est observé dans le cas des protéines complètes. Ceci suggère que la portion UP1 contenant les 2 domaines de liaison à l'ARN est importante pour permettre des interactions fortes entre les molécules hnRNP A1, possiblement en permettant de stabiliser l'interaction entre les protéines.

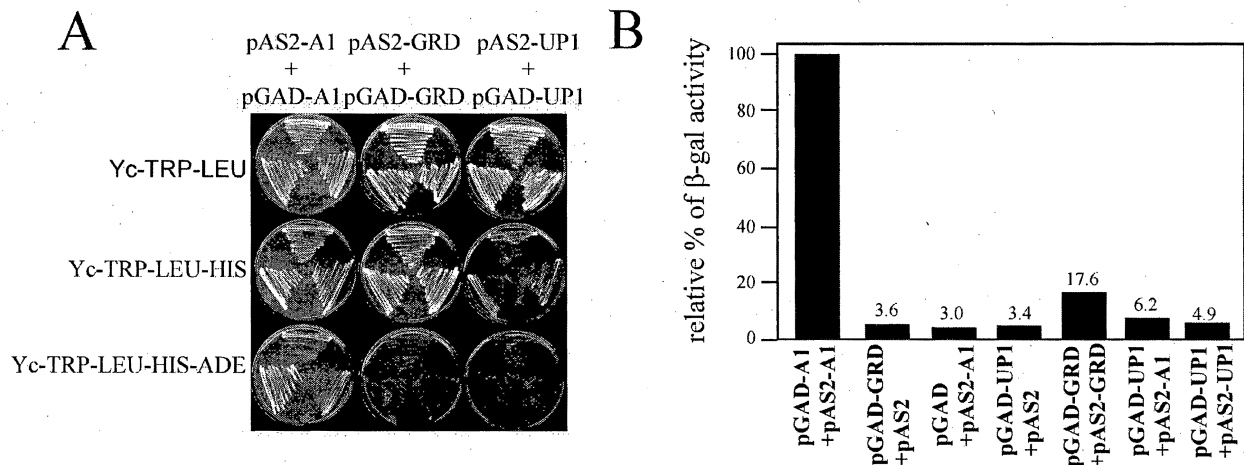


Figure 21. Caractérisation des interactions entre les domaines de la protéine A1 par l'utilisation du système double-hybride. A) Trois clones individuels résultant des transformations avec les différents plasmides de double-hybride contenant les inserts indiqués ont été striés sur les milieux de sélection mentionnés, et leur croissance a été analysée. Les contrôles appropriés ont aussi été analysés (données non-montrées). B) Les souches contenant les différentes combinaisons de vecteurs ont été analysées pour l'activité relative de β-Galactosidase. Les niveaux relatifs sont comparés à une interaction entre deux protéines hnRNP A1 complètes, qui donne un pourcentage relatif de 100% pour l'activité β-Galactosidase. Chacune des différentes combinaisons représente une compilation de 5 clones indépendants pour l'activité β-Galactosidase.

La deuxième approche utilisée a été d'introduire des mutations ciblées dans la protéine hnRNP A1. Le choix des différents résidus à mutagéniser de façon dirigée a été fait selon la structure du cristal d'UP1 avec un oligonucléotide comportant deux répétitions télomériques humaines (Ding et al. 1999) (figure 22). L'idée générale est de

cibler les acides aminés importants dans l'interaction entre UP1 et un oligo télomérique et entre les molécules UP1 elles-mêmes. En apportant des changements à ces résidus, on pourrait voir si l'interaction entre protéines est dépendante de la liaison à un acide nucléique.

Des mutations ponctuelles ont été introduites dans l'ADNc de hnRNP A1 par une méthode d'extension d'amorces qui se chevauchent (overlap extension). Certains acides aminés ont été ciblés pour leur importance dans l'interaction entre deux molécules d'UP1 (i.e. glutamate 11; E11A, aspartate 94; D94A, histidine 173; H173A, et lysine 166; K166A, en vert sur la figure 22). Un autre résidu ciblé semble important dans la liaison des acides nucléiques (méthionine 46; M46V, en mauve sur la figure 22). Le dernier résidu ciblé est dans une région quelconque afin de créer un contrôle pour les autres mutations ciblées (valine 44; V44A).

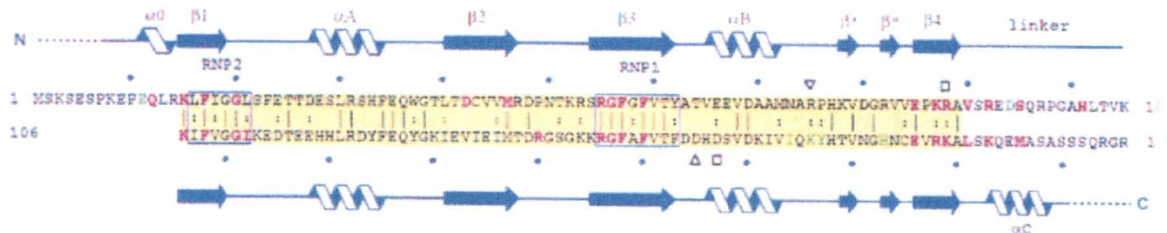


Figure 22. Séquence des acides aminés de UP1 et structure de la protéine. Les deux RRM sont dans des boîtes jaunes. Les motifs RNP1 et RNP2 sont mentionnés. Le numéro des résidus est indiqué au début de chaque ligne, et un point noir indique une augmentation de 10 dans le numéro des résidus. La structure secondaire des éléments est indiquée au-dessus et au-dessous de la séquence. Les résidus en vert sont impliqués dans des interactions protéine : protéine entre les monomères. Les résidus indiqués en mauve sont impliqués dans la liaison de l'oligo télomérique simple-brin. Tiré de Ding et al. (1999)

Par la suite, certaines de ces mutations ont été combinées entre elles pour créer des protéines comportant deux ou même trois mutations. Les protéines recombinantes GST-A1 comportant ces mutations ont été produites et séquencées, à partir du vecteur d'expression pGEX. Ces protéines GST-A1 mutantes ont été analysées pour leur capacité

à lier un ARN contenant une séquence de liaison de haute affinité pour la protéine de type sauvage (CE4). Par la suite, leur habileté à moduler la sélection des sites d'épissage alternatif sur un transcrit modèle *in vitro* a été vérifiée. Aussi, suite au clonage des ADNc contenant les mutations dans les vecteurs d'expression pour la levure, la capacité d'interaction des différentes protéines mutantes a été analysée par double-hybride *in vivo*.

Le tableau 5 résume les résultats obtenus avec les différentes protéines contenant des mutations ciblées dans le domaine UP1. Malgré les différentes mutations et combinaisons de mutations insérées dans l'ADNc de hnRNP A1, aucune différence n'est perçue entre les protéines mutantes et les protéines de type sauvage, pour les 3 différents aspects étudiés. Les différentes mutations ciblées ne créent aucun effet significatif dans les fonctions de la protéine hnRNP A1. Les mutations insérées dans la protéine A1 étaient basées sur la structure du cristal de UP1. Elles consistaient en différents essais pour voir si le changement d'un résidu positionné à un endroit stratégique dans la protéine, permettait de changer le mode d'action de la protéine complète. Malheureusement, aucune mutation ni combinaison de mutations n'a donné comme résultat les caractéristiques recherchées pour une perte de fonction de la protéine A1.

Mutation ciblée dans la protéine	Liaison à l'ARN	Modulation de l'épissage d'un transcrit modèle	Interaction protéine-protéine (double-hybride)
hnRNP A1	+++	distal	+++
hnRNP A1 M46V	+++	distal	+++
hnRNP A1 D94A	+++	distal	+++
hnRNP A1 M46A+D94A	+++	distal	++(+)
hnRNP A1 V44A	++(+)	distal	+++
hnRNP A1 E11A+K166A	+++	distal	+++
hnRNP A1 E11A+K166A+H173A	+++	distal	+++

Tableau 5. Résumé des résultats obtenus avec les protéines hnRNP A1 contenant des mutations à des résidus importants pour certaines interactions des protéines de type sauvage. Les mutations ont été introduites dans l'ADNc de

hnRNP A1 et les essais ont été faits avec des protéines recombinantes GST-A1 mutantes (liaison à l'ARN et modulation de l'épissage). L'ARN utilisé pour les essais de liaison est un oligo d'ARN CE4, dérivé de la séquence intronique de hnRNP A1. Environ 0.15 µg, 0.45 µg et 1.25 µg de protéines recombinantes ont été testées pour la liaison à l'ARN et analysées sur un gel natif 6% polyacrylamide. Les protéines recombinantes GST-A1 mutantes ont aussi été testées dans des essais d'épissage *in vitro* sur un transcrit modèle contenant des sites de liaison pour hnRNP A1 de part et d'autre de l'exon alternatif (transcrit 86.6). Les extraits nucléaires utilisés ont été dépourvus des protéines endogènes de la famille hnRNP A/B par passage sur une résine à laquelle est couplé un oligo d'ARN. Les différentes réactions d'épissage ont été faites avec 0.28 µg, 0.83 µg et 2.5 µg de protéines recombinantes, à 30°C pour 2 heures, et mises sur un gel 11% polyacrylamide dénaturant. Chaque protéine mutante a été analysée 3 fois pour sa modulation de l'épissage alternatif. L'ADNc de chaque mutant a été inséré dans le plasmide pAS2-1A et les interactions ont été vérifiées entre une protéine hnRNP A1 mutante et une protéine hnRNP A1 de type sauvage. Les tests d'interaction ont été faits par activation des gènes rapporteurs, et répétés 3 fois pour chacune des protéines mutantes.

Comme l'approche d'insertion de mutations ciblées dans la protéine hnRNP A1 s'est avérée négative, une autre approche a été envisagée. Dans celle-ci, les mutations introduites dans la protéine A1 étaient décidées par la nature, dans le but de trouver des positions dans la protéine ayant une importance significative pour sa fonction. Nous avons entrepris une approche de mutation aléatoire de l'ADNc de hnRNP A1, en utilisant des conditions de PCR qui augmentent le nombre d'erreurs introduites par la Taq polymérase lors de l'amplification d'un gène. Le but visé est de récupérer des protéines hnRNP A1 mutantes, qui sont complètes et qui ont un défaut au niveau de l'interaction protéine : protéine. Ensuite, il s'agit de vérifier l'effet de la perte d'interaction entre des protéines A1 complètes face à la modulation de l'épissage alternatif.

Les conditions utilisées pour la mutagenèse permettent d'introduire environ 1 ou 2 mutations par 1000 pb (Gravel et Wellinger 2002). Comme le gène de hnRNP A1 est d'environ 1.4 kb, il est possible d'obtenir de 1 à 3 mutations par ADNc ayant été amplifié selon ces conditions. Les amorces pour ces réactions ont été choisies afin de conserver des régions d'homologie en amont et en aval de l'ADNc. Ces régions se retrouvent également présentes dans le plasmide de double-hybride utilisé, et serviront ensuite dans l'étape de clonage via recombinaison homologue (figure 23). Le fragment d'ADNc portant des mutations suite au protocole d'amplification par PCR a été purifié. Par la suite, il a été co-transformé dans la souche de levure PJ69-4a Ku80::ura3 + pGAD-A1-8,

en même temps qu'un plasmide pAS2-1A linéarisé par digestion au site unique BamHI. Un protocole de transformation à haute efficacité a été utilisé pour maximiser l'entrée des fragments dans la souche de levure. Étant donné que les fragments linéaires sans télomères ne sont pas maintenus dans la levure, il doit y avoir recombinaison homologue entre les fragments pour créer un plasmide circulaire. Donc, il peut y avoir « appariement ou reconnaissance » entre les séquences homologues du plasmide linéarisé et celles des fragments d'ADNc amplifiés par PCR (situées en 5' et en 3' du fragment amplifié). Ces séquences homologues permettent d'enclencher un processus complexe de recombinaison homologue (revue dans Paques et Haber 1999). La délétion endogène du gène Ku80 dans la souche de levure permet d'augmenter la fréquence des événements de recombinaison homologue. En effet, la majorité des événements de réparation des cassures d'ADN dans la levure sont des événements NHEJ (non-homologous end-joining) (revue dans Paques et Haber 1999). Comme le complexe protéique Ku a un rôle majeur à jouer dans ce processus de NHEJ (Tsukamoto et al. 1996; Boulton et Jackson 1996), la délétion d'une sous-unité de ce complexe permet de le rendre inactif et permet de diminuer la fréquence du NHEJ. Le produit final du processus de recombinaison homologue dans la levure est un plasmide circulaire. Celui-ci contient le fragment d'ADNc comportant les mutations introduites par PCR et la base du vecteur pAS2-1. Il permet la production d'une protéine de fusion GAL4 DBD-A1 mutante, ainsi que la réplication et la sélection du plasmide (figure 23).

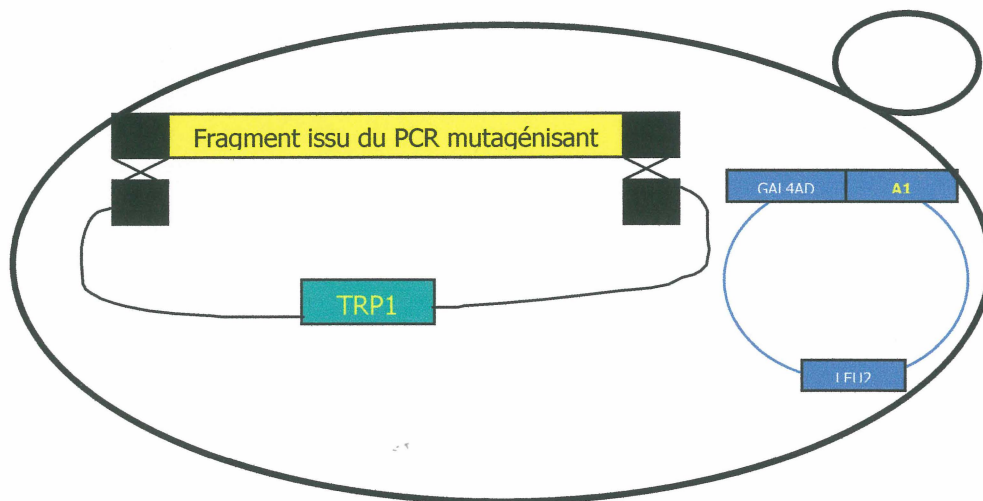


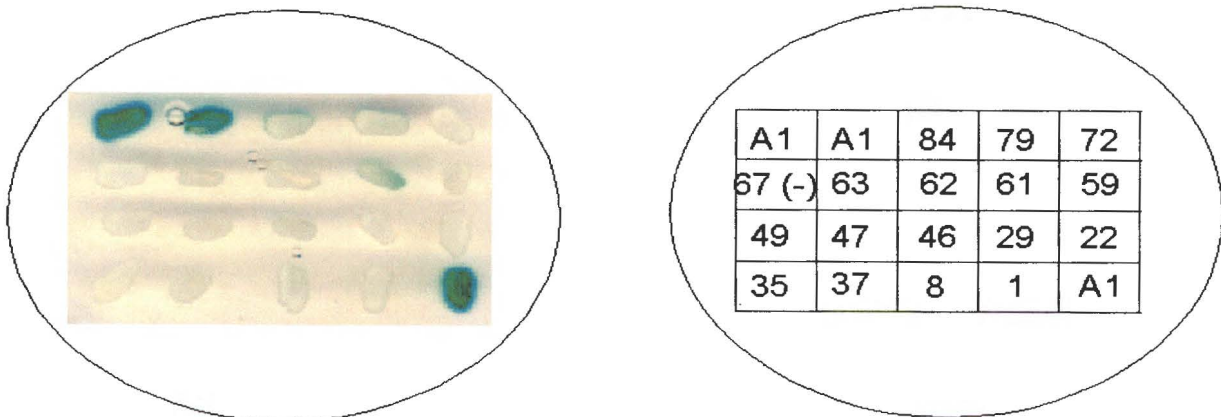
Figure 23. Aperçu de la recombinaison homologue menant à un plasmide linéaire. La souche de levure utilisée au départ contient le plasmide GAL4 AD-A1 (LEU2) (bleu). Cette souche subit la co-transformation de 2 fragments (le fragment d'ADNc issu du PCR et le plasmide pAS2-1 linéarisé). Le fragment linéaire contient le marqueur de sélection TRP1 (vert), et des séquences homologues à celles retrouvées en 3' et 5' de l'ADNc de A1 (boîtes noires). Le fragment issu de la réaction de PCR contient le gène de A1 avec certaines mutations incorporées dans la séquence (jaune), et des séquences homologues en 3' et en 5' à celles retrouvées sur le plasmide linéarisé (boîtes noires). Les deux fragments peuvent recombiner entre eux et mener à la formation d'un plasmide circulaire. La souche de levure résultante peut être testée pour les interactions entre une protéine GAL4 AD-A1 et une protéine GAL4 DBD-A1 mutante.

En étalant le résultat de la co-transformation sur des boîtes Yc-Trp-Leu, nous avons obtenu des colonies dans lesquelles le plasmide pAS2-1A a été circularisé grâce à l'intégration du fragment d'ADNc de A1 contenant les mutations. Par la suite, les interactions entre les protéines GAL4 DBD-A1 mutante et GAL4 AD-A1 contenues dans ces levures ont été testées, par essai β -Gal sur boîte. Les colonies qui sont demeurées blanches suite à cet essai ont été récupérées, car ce sont elles qui montrent une perte d'interaction entre les protéines. L'étape suivante est de s'assurer que la perte d'interaction détectée est dépendante des mutations introduites aléatoirement dans la protéine hnRNP A1 complète. En effet, plusieurs autres conditions peuvent mener à une perte d'interaction entre les protéines testées. Par exemple, il peut y avoir 1) recircularisation du plasmide par NHEJ, ou 2) recombinaison entre les séquences homologues du vecteur linéaire et du vecteur circulaire déjà présent dans la souche de levure ou encore 3) introduction d'un codon STOP à la fin du domaine de liaison à

l'ADN de GAL4 ou de façon précoce dans l'ADNc de hnRNP A1. Ces différents exemples résultent en un phénotype de perte d'interaction entre les protéines, mais sont considérés comme des faux positifs car la protéine A1 mutante n'est pas complète. Pour savoir si la protéine A1 mutante est complète dans la levure, des extraits de protéines ont été faits à partir des différentes colonies. Les protéines de fusion GAL4 DBD-A1 mutante ont été analysées par Western Blot avec un anticorps anti-GAL4 DBD. Seules les colonies qui possédaient des protéines de fusion GAL4 DBD-A1 complètes ont été conservées.

Environ 2000 colonies ont été récupérées suite à la co-transformation et à la sélection sur un milieu Yc-Trp-Leu. Seulement 244 de ces colonies ont donné des colonies blanches suite à l'essai β -Gal sur boîte (figure 24A). Ces colonies ont également été striées sur des milieux Yc-Trp-Leu-His-Ade, pour vérifier la perte d'interaction sur les milieux de sélection. Suite à l'analyse des protéines de fusion par Western Blot, seulement 28 colonies contenaient une protéine de fusion GAL4 DBD-A1 mutante complète et 9 contenaient une forme tronquée (figure 24B).

A)



B)

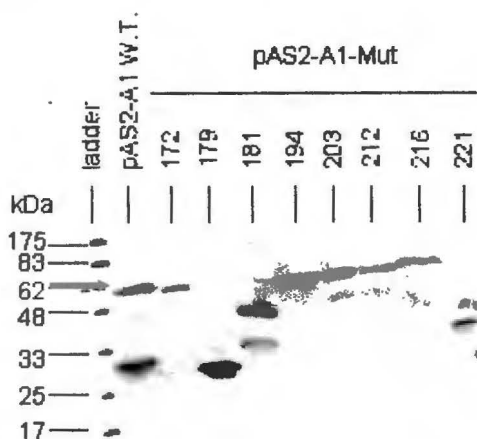


Figure 24. Analyses de la perte d'interaction entre les protéines hnRNP A1 contenant des mutations aléatoires. A) Gauche : Essai β -Gal sur boîte avec des levures contenant les deux plasmides circulaires pAS2-A1 mutant et pGAD-A1. Droite : Identification des différents clones du panneau de gauche. La mention A1 réfère à une interaction entre deux protéines de type sauvage, d'où la couleur bleue qui est observée. La mention 67(-) correspond à une souche de levure

ne comprenant pas la version pGAD-A1. Les essais ont été faits en triplicata (trois boîtes différentes contenant les mêmes souches de levure). Les résultats colorimétriques obtenus concernant l'interaction entre les protéines mutantes ont été comparés. Les colonies blanches récupérées ont été analysées pour la présence du plasmide pAS2-A1 encodant une protéine de fusion GAL4 DBD-A1 mutante complète. B) Analyse Western Blot des protéines de fusion GAL4 DBD-A1

mutante. Les protéines ont été isolées des cultures de levures puis un aliquot de l'extrait a été chargé sur gel. Suite au transfert sur membrane de nitrocellulose, il y a eu hybridation avec un anticorps primaire anti-GAL4 DBD. L'anticorps secondaire (anti-souris) et la réaction de chemiluminescence ont permis de visualiser les protéines sur autoradiogramme. Les protéines de fusion complètes sont d'environ 58 kDa (flèche rouge). Les bandes plus courtes (échantillons 179, 181, 221) peuvent correspondre à des formes tronquées de la protéine de fusion ou à une protéolyse de la protéine dans les extraits.

Dix-huit différents ADNc encodant des protéines hnRNP A1 mutantes complètes ont été clonés dans le vecteur pGEX-KG. Les plasmides obtenus ont été séquencés afin de s'assurer du respect du cadre de lecture et de localiser les positions des différentes mutations. Les protéines recombinantes GST-A1 mutantes ont été produites dans des bactéries et vérifiées par Western Blot (anticorps primaire α -A1/UP1). Par la suite ces protéines GST-A1 mutantes ont été analysées pour leur capacité à lier un substrat ARN et leur capacité à moduler l'épissage alternatif *in vitro* sur un transcrit modèle.

Le tableau 6 dresse un résumé des résultats obtenus avec les protéines A1 comportant une ou des mutations aléatoires menant à une perte d'interaction. Certaines des protéines GST-A1 mutantes ont perdu leur capacité à lier un substrat ARN contenant un site de liaison de haute affinité pour A1. Ces protéines n'ont par la suite pas été étudiées pour leur capacité de moduler l'épissage alternatif. Certaines autres protéines GST-A1 mutantes ont gardé leur capacité à lier l'ARN et aussi à moduler l'épissage

alternatif d'un transcrit modèle *in vitro*, dans des extraits dépourvus de protéines hnRNP A/B endogènes. Quand on compare les différentes mutations récupérées suite à cette procédure, certaines de ces mutations sont retrouvées plus d'une fois dans les protéines mutantes.

Protéine GST-A1	Mutation(s) présente(s)	Localisation de la mutation dans la protéine	Liaison à l'ARN (CE4)	Modulation de l'épissage	Interaction protéine-protéine (LacZ sur boîte)	Interaction protéine-protéine (LacZ liquide)
GST-A1	-----	-----	+++	distal	Bleu	100%
16	Stop AA 276	GRD	++	n/d	Blanc	2.1%
29	??	??	-----	n/d	Blanc	1.5%
37	M72V	RRM1	+++	distal	Bleu pâle	9.3%
46	V151D	RRM2	-----	n/d	Blanc	3.2%
47	Q165K + Stop AA177	RRM2 + stop	-----	n/d	Blanc	2.2%
61	M72V	RRM1	++	distal	Bleu pâle	12.5%
63	F153L	RRM2	+++	distal	Bleu très pâle	6.8%
72	??	??	+++	n/d	Blanc	2.0%
79	F125S + Q165K	RRM2 + RRM2	----	n/d	Bleu pâle	6.5%
84	K179E	RRM2	+++	n/d	Bleu pâle	1.44%
102	V159D + L181Q	RRM2 + protéine	+++	distal	Blanc	1.1%
107	I107Y + Q165K	RRM2 + RRM2	----	majorité proximal	blanc	2.5%
144	L21P	RRM1	----	n/d	Bleu pâle	11.5%
145	V151D	RRM2	----	n/d	Blanc	2.6%
153	G58R	RRM1	----	n/d	Blanc	2.4%
203	H77R	RRM1	++	n/d	Blanc	2.6%
212	F153I + V170G	RRM2 + RRM2	---	n/d	Bleu pâle	3.9%
59	Q165? + Q193?	RRM2	+++	n/d	Bleu pâle	12.2%

Tableau 6. Résumé des données récupérées suite aux analyses faites sur les protéines de fusion GST-A1 avec des mutations aléatoires. Les protéines de fusion GST-A1 mutantes ont été séquencées pour trouver la (les) mutation(s) responsable(s) de la perte d'interaction entre les protéines. Ces protéines de fusion ont été analysées pour leur liaison à l'ARN CE4, et certaines d'entre elles ont été étudiées pour leur capacité à moduler l'épissage alternatif d'un transcrit modèle *in vitro*. Les différentes protéines ont aussi été analysées pour leur interaction, par essai colorimétrique pour obtenir une activité relative de β -Galactosidase. Les niveaux relatifs sont comparés à une interaction entre deux protéines hnRNP A1 de type sauvage (pourcentage relatif de 100% pour l'activité β -Galactosidase). Chacune des

différentes combinaisons représente une compilation de 3 expériences indépendantes pour l'activité β -Galactosidase. n/d = non déterminé.

Ces différents sites plus sensibles à la mutagenèse peuvent possiblement indiquer un endroit important pour l'interaction entre les protéines (par exemple, la mutation en position 72 retrouvée dans les clones GST-A1-37 et GST-A1-61, ou la mutation en position 151 retrouvée dans les clones GST-A1-46 et GST-A1-145; tableau 6).

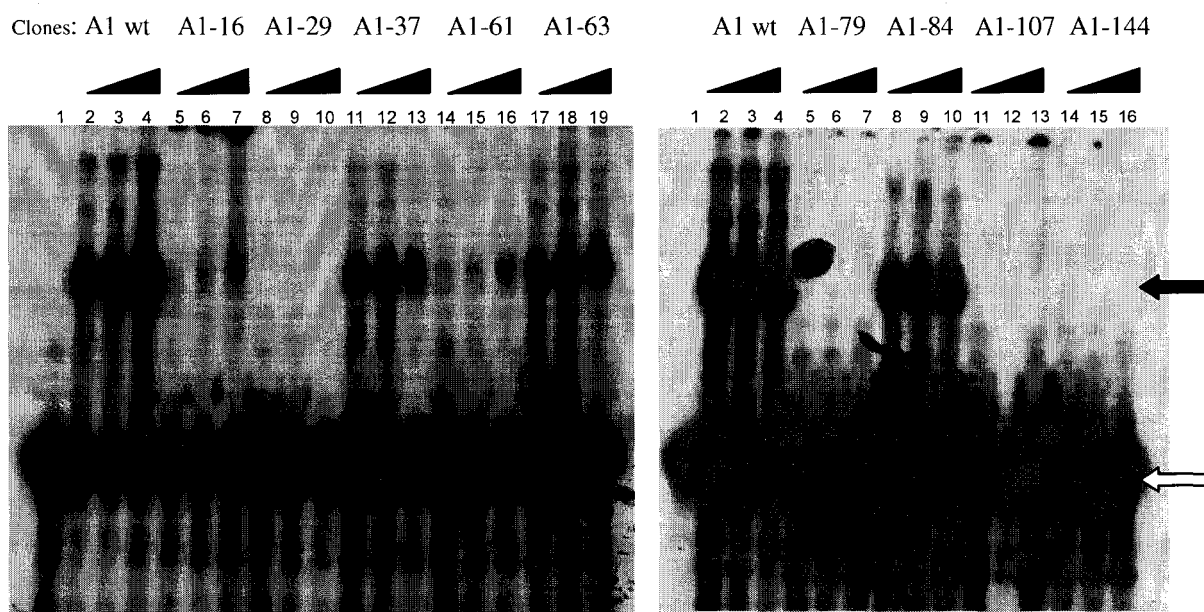


Figure 25. Gels de retardement des protéines recombinantes GST-A1 mutantes avec un oligo ARN.

Les protéines recombinantes produites ont été dosées pour avoir 0,14 µg, 0,42 µg et 1,25 µg de protéines. La liaison avec le substrat ARN CE4 a été effectuée à 4° C pour 5 minutes. Les complexes ont été chargés sur un gel 6 % polyacrylamide natif, migré à 150 V pour 3 heures 30 minutes. Le gel a été séché et exposé sur autoradiogramme.

Gauche, pistes : 1) aucune protéine 2) GST-A1 (0,14 µg) 3) GST-A1 (0,42 µg) 4) GST-A1 (1,25 µg) 5) GST-A1-16 (0,14 µg) 6) GST-A1-16 (0,42 µg) 7) GST-A1-16 (1,25 µg) 8) GST-A1-29 (0,14 µg) 9) GST-A1-29 (0,42 µg) 10) GST-A1-29 (1,25 µg) 11) GST-A1-37 (0,14 µg) 12) GST-A1-37 (0,42 µg) 13) GST-A1-37 (1,25 µg) 14) GST-A1-61 (0,14 µg) 15) GST-A1-61 (0,42 µg) 16) GST-A1-61 (1,25 µg) 17) GST-A1-63 (0,14 µg) 18) GST-A1-63 (0,42 µg) 19) GST-A1-63 (1,25 µg).

Droite, pistes : 1) aucune protéine 2) GST-A1 (0,14 µg) 3) GST-A1 (0,42 µg) 4) GST-A1 (1,25 µg) 5) GST-A1-79 (0,14 µg) 6) GST-A1-79 (0,42 µg) 7) GST-A1-79 (1,25 µg) 8) GST-A1-84 (0,14 µg) 9) GST-A1-84 (0,42 µg) 10) GST-A1-84 (1,25 µg) 11) GST-A1-107 (0,14 µg) 12) GST-A1-107 (0,42 µg) 13) GST-A1-107 (1,25 µg) 14) GST-A1-144 (0,14 µg) 15) GST-A1-144 (0,42 µg) 16) GST-A1-144 (1,25 µg). La flèche noire indique la position d'un complexe entre l'oligo radiomarqué CE4 et les protéines recombinantes. La flèche blanche indique l'oligo radiomarqué CE4 libre.

Des mutations retrouvées dans les protéines GST-A1 mutantes (tableau 6 et figure 25) ont mené à une diminution de leur capacité d'interaction avec un substrat ARN

(protéines GST-A1-16, GST-A1-61, GST-A1-203) ou à une perte totale de leur capacité d'interaction avec un substrat ARN (protéines GST-A1-29, GST-A1-46, GST-A1-47, GST-A1-79, GST-A1-107, GST-A1-144, 145, GST-A1-153, GST-A1-212).

Dans le cas des protéines GST-A1-144 et GST-A1-153, une seule mutation d'un résidu situé dans le RRM1 (L21P et G58R, respectivement) mène à une perte totale de la capacité de liaison à l'ARN. Cependant, les protéines GST-A1-203 et GST-A1-61 possèdent aussi une seule mutation dans le RRM1 (H77R et M72V, respectivement) et peuvent encore interagir avec le substrat ARN, bien que de façon moins efficace que la protéine de type sauvage. La protéine GST-A1-37 contient elle aussi la mutation M72V dans le RRM1 et peut cependant interagir avec le substrat ARN presque aussi efficacement que la protéine GST-A1 (voir plus loin). L'introduction d'un codon STOP en position 276 dans la protéine GST-A1-16 mène à une diminution de sa capacité d'interaction avec le substrat ARN, même si aucune mutation n'est présente dans les RRM.

Dans le cas des protéines GST-A1-46, GST-A1-47, GST-A1-79, GST-A1-107, GST-A1-145, GST-A1-212, des mutations dans le RRM2 causent une perte totale de leur capacité à lier l'ARN. Dans le cas des autres protéines possédant des mutations dans le RRM2 (GST-A1-63, GST-A1-84, GST-A1-102, et GST-A1-59) les protéines demeurent capables de lier l'ARN, malgré leurs mutations.

Suite à l'analyse de ces données, il semble que les mutations de différents résidus aient des conséquences variables pour la protéine. Dans le RRM1, les mutations qui abolissent l'interaction avec l'ARN sont localisées dans les domaines RNPs des RRM, qui sont les motifs directement impliqués dans la liaison à l'ARN (GST-A1-144 et GST-

A1-153). Même si les résidus mutés ne sont pas directement en contact avec l'ARN (voir figure 22), il semble tout de même que ces résidus soient impliqués dans la formation d'une structure globale qui permet la liaison au substrat, et leur mutation a des conséquences graves pour cette liaison.

Dans deux cas, la mutation du résidu 72 (M72V) situé dans le RRM1 est retrouvée. Ce changement mène au remplacement d'un résidu hydrophobe contenant un atome de soufre par un résidu hydrophobe à courte chaîne. Dans la protéine GST-A1-37, la mutation ne semble pas affecter la capacité de liaison à l'ARN de la protéine, mais pour la protéine GST-A1-61, la mutation diminue l'efficacité de liaison à l'ARN. Cependant, la protéine GST-A1-203 possède aussi seulement une mutation dans la portion terminale du RRM1 (H77R, résidu basique avec un cycle changé par un résidu basique à longue chaîne latérale), et sa capacité à lier l'ARN est diminuée, comme pour la protéine GST-A1-61. La présence d'une seconde mutation non-détectée dans la protéine GST-A1-61 peut être une explication pour la différence observée entre les protéines dans la capacité d'interaction avec l'ARN.

Les mutations affectant le domaine RRM2 semblent avoir des conséquences plus complexes. Encore une fois, certaines mutations retrouvées plus d'une fois ont des conséquences variables pour les protéines produites, par exemple GST-A1-63 (F153L, résidu à cycle aromatique muté pour un résidu à chaîne latérale simple) et GST-A1-212 (F153I + V170G, résidu à chaîne aromatique pour résidu à chaîne latérale et résidu à chaîne latérale pour résidu simple, respectivement). Dans ce cas, GST-A1-63 peut lier l'ARN et GST-A1-212 en est incapable. Selon cette comparaison, la perte de la capacité à lier l'ARN serait attribuée à la mutation du résidu 170 (situé dans la portion terminale du

RRM2). Par contre, le résidu 153 (localisé dans le domaine RNP1 du RRM2) devrait être celui dont la mutation a un effet sur la liaison à l'ARN (figure 22). Peut-être que la mutation de ce résidu phénylalanine en leucine permet de garder une conformation adéquate pour la liaison à l'ARN (GST-A1-63) alors que sa mutation en isoleucine ne le permet pas (GST-A1 212).

D'autres mutations récupérées dans le RNP1 du RRM2 mènent à une perte de l'interaction avec l'ARN (GST-A1-145 et GST-A1-46). Comme dans le cas des mutations dans les RNPs du RRM1, ces résidus ne sont pas identifiés comme interagissant directement avec les acides nucléiques, mais leur mutation mène à une perte de la liaison à l'ARN. Par contre, la protéine GST-A1-84 comporte la mutation dans un résidu (K179E, un résidu basique pour un résidu acide) qui est impliqué dans les contacts avec l'acide nucléique dans la portion terminale du RRM2 et sa mutation n'affecte pas la liaison à l'ARN. Aussi, la protéine GST-A1-102 comporte deux mutations dans la portion terminale du RRM2 (V159D, un résidu hydrophobe non-polaire pour un résidu hydrophile polaire et basique, et L181Q, un résidu hydrophobe non-polaire pour un résidu hydrophile polaire). La présence de ces deux mutations, malgré le fait que la position 181 semble impliquée dans la liaison de l'acide nucléique, ne réduit pas la liaison de la protéine à l'ARN.

Une autre mutation (Q165K) est retrouvée dans plusieurs protéines mutantes (GST-A1-47, GST-A1-79, GST-A1-107, GST-A1-59). Dans cette mutation, un résidu neutre est changé pour un résidu basique. Ce résidu se retrouve dans une région du RRM2 où sont concentrés les résidus importants dans l'interaction protéine : protéine. Cette mutation peut être à l'origine de la perte d'interaction entre les protéines, mais les

autres mutations dans les protéines devraient être celles qui ont de l'importance quant à la liaison à l'acide nucléique. Donc, les mutations F125S et I107Y retrouvées dans les protéines GST-A1-79 et GST-A1-107, respectivement, seraient responsables de la perte de la liaison à l'ARN. En effet, la position 107 est située dans le RNP2 du RRM2, donc une mutation à cet endroit peut être la cause de la perte de l'interaction avec l'ARN. Par contre, la position 125 ne se situe ni dans le RNP2 ni dans le RNP1 du RRM2 et mène aussi à une perte de liaison de l'ARN. Ceci indiquerait que d'autres résidus situés en dehors de RNPs ont des conséquences importantes sur la liaison à l'ARN ou alors dans ce cas, c'est la combinaison des deux mutations qui mène à cet effet. Le cas de la protéine GST-A1-107 est à part car elle contient aussi un codon STOP en position 177, ce qui en fait une protéine tronquée.

Certaines protéines mutantes ont été étudiées dans des essais d'épissage alternatif *in vitro* (figure 26). Ces essais sont faits dans des extraits nucléaires qui sont dépourvus des protéines hnRNP A1/A1^B/A2/B1, suite à leur passage sur une résine. Cette résine comporte un court ARN CE1a qui lui est couplé, et sur lequel les protéines vont se lier. Ces extraits sont vérifiés par Western Blot (donnée non-montrée) afin de s'assurer que les protéines endogènes ont bien été enlevées de l'extrait nucléaire. L'action des différentes protéines recombinantes sur la sélection des sites d'épissage d'un transcrit modèle est analysée. Ceci est possible grâce au rajout en quantité croissante des protéines recombinantes dans la réaction d'épissage qui est faite dans un extrait dépourvu de protéines A1. L'épissage dans un extrait nucléaire dépourvu est moins efficace que l'épissage dans un extrait normal, à cause de la manipulation de l'extrait et du passage sur la résine. L'épissage d'un transcrit modèle en présence de la protéine A1 permet de

favoriser l'utilisation d'un site d'épissage distal, donc la production de produits distaux (figure 26, flèche rouge) (Mayeda et Krainer 1992; Mayeda et al. 1993; Mayeda et al. 1994; Caceres et al. 1994; Yang et al. 1994; Chabot et al. 1997; Blanchette et Chabot 1999; Hutchison et al. 2002; Nasim et al. 2002). L'épissage d'un transcrit modèle dans un extrait nucléaire dépourvu de protéines A1 montre une production majeure de produits proximaux, dus à l'utilisation du site d'épissage proximal (figure 26, flèche noire).

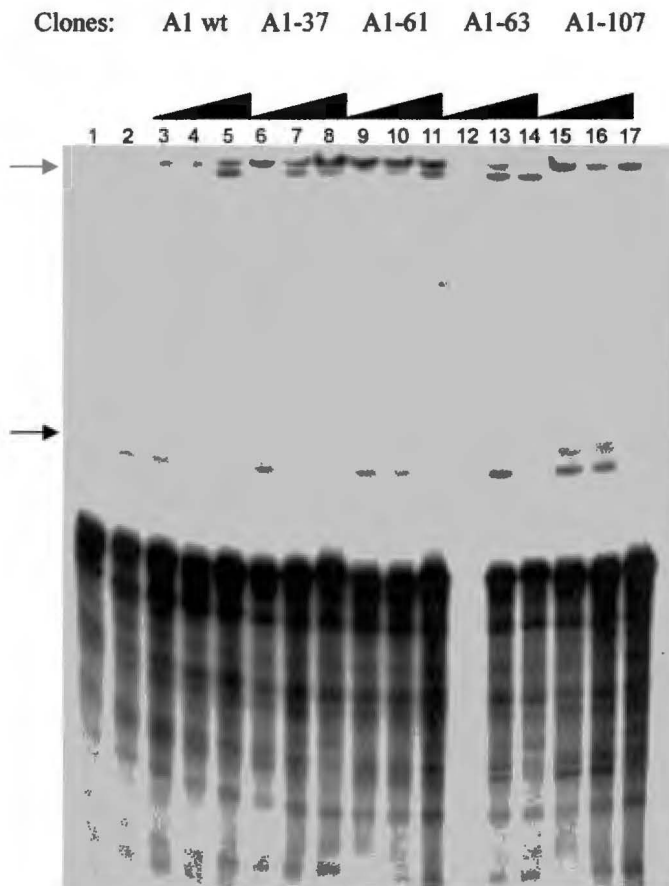


Figure 26. Épissage *in vitro* avec les protéines GST-A1 mutantes. Les extraits nucléaires de cellules HeLa ont été mis sur une résine à laquelle est un couplé un oligo ARN (CE1a), ce qui permet la déplétion des protéines endogènes hnRNPs A/B (extrait nucléaire dépourvu). Le transcrit 86.6 contenant deux sites de liaison pour A1 est utilisé pour l'épissage dans ces extraits. Les protéines recombinantes GST-A1 (0,3 µg, 0,9 µg et 2,5 µg) sont rajoutées dans la réaction d'épissage qui est incubée à 30 °C pour 2 heures. Les produits d'épissage sont chargés sur un gel polyacrylamide dénaturant 11%. Pistes : 1) extrait nucléaire, mis sur résine 2) extrait nucléaire dépourvu de A1 3) extrait nucléaire dépourvu + GST-A1 (0,3 µg) 4) extrait nucléaire dépourvu + GST-A1 (0,9 µg) 5) extrait nucléaire dépourvu + GST-A1 (2,5 µg) 6) extrait nucléaire dépourvu + GST-A1-37 (0,3 µg) 7) extrait nucléaire dépourvu + GST-A1-37 (0,9 µg) 8) extrait nucléaire dépourvu + GST-A1-37 (2,5 µg) 9) extrait nucléaire dépourvu + GST-A1-61 (0,3 µg) 10) extrait nucléaire dépourvu + GST-A1-61 (0,9 µg) 11) extrait nucléaire dépourvu + GST-A1-61 (2,5 µg) 12) extrait nucléaire dépourvu + GST-A1-63 (0,3 µg) 13) extrait nucléaire dépourvu + GST-A1-63 (0,9 µg) 14) extrait nucléaire dépourvu + GST-A1-63 (2,5 µg) 15) extrait nucléaire dépourvu + GST-A1-107 (0,3 µg) 16) extrait nucléaire dépourvu + GST-A1-107 (0,9 µg) 17) extrait nucléaire dépourvu + GST-A1-107 (2,5 µg). La flèche rouge indique la position des produits d'épissage distaux et la flèche noire indique la position des produits d'épissage proximaux.

Pour être actives dans l'épissage, les protéines doivent être capables de lier l'ARN. Pour ce faire, les protéines GST-A1-37, GST-A1-61, GST-A1-63 et GST-A1-107 (contrôle négatif) ont été testées dans l'épissage d'un transcrit modèle 86.6. Ce transcrit

contient des sites de liaison de haute affinité pour A1 de part et d'autre de l'exon alternatif. Sur la figure 26, l'extrait nucléaire passé sur la résine seule (piste 1) contient encore les protéines A1, et montre un ratio de produits d'épissage contenant plus de produits distaux (flèche rouge) que proximaux (flèche noire). L'extrait passé sur une résine couplée à un oligo CE1a (extrait dépourvu des protéines endogènes) montre le patron d'épissage d'un extrait ne contenant plus de protéines A1, soit aucun produit distal et un produit proximal majoritaire (piste 2). L'épissage dans un extrait nucléaire dépourvu de A1 dans lequel on rajoute la protéine A1 recombinante montre un patron d'épissage dans lequel la production des produits distaux est restaurée, avec la concentration croissante de protéines A1 rajoutée (pistes 3-4-5).

Des protéines qui ont perdu leur capacité d'interagir entre elles mais qui ont gardé leur capacité de lier l'ARN ont été testées dans des essais d'épissage *in vitro*. Selon le modèle postulé, l'interaction entre les protéines A1 permet de rapprocher des sites d'épissage distancés pour permettre la sélection d'un site d'épissage distal (Blanchette and Chabot 1999). Dans ce cas, l'épissage d'un transcrit modèle dans un extrait dépourvu en A1 dans lequel on rajoute des protéines incapables d'interagir entre elles devrait donner un patron d'épissage semblable à celui observé dans un extrait dépourvu de A1 (piste 2), soit une majorité de produit proximal. Dans les cas des protéines GST-A1-37, GST-A1-61 et GST-A1-63, le patron d'épissage alternatif observé est semblable à celui observé dans un extrait nucléaire normal, soit l'augmentation des produits proximaux en fonction de la concentration de protéine recombinante rajoutée (pistes 6-7-8, 9-10-11, 12-13-14). La protéine GST-A1-107 ayant perdu sa capacité de lier l'ARN montre un patron d'épissage d'un extrait dépourvu de A1. Une très faible proportion de produits distaux est

observée en présence d'une forte concentration de protéines recombinantes (pistes 15-16-17). Il a été observé auparavant qu'une trop grande concentration de protéines recombinantes peut causer un épissage aberrant d'un transcrit. Donc, les protéines mutantes incapables d'interagir entre elles qui ont été testées dans des essais d'épissage *in vitro* semblent pouvoir modifier l'épissage alternatif du transcrit modèle comme une protéine A1 de type sauvage. Ceci peut suggérer que l'interaction entre les protéines A1 a peu d'importance pour la modulation de l'épissage alternatif. Encore, il se peut que l'interaction entre les protéines GST-A1-mutantes liées à l'ARN soit permise et accomplie par le domaine GST des protéines recombinantes. En effet, il a été observé auparavant que ces domaines ont une affinité et une capacité de liaison entre eux (Niedziela-Majka et al. 1998). Étant donné que la protéine GST-A1-107 n'est pas liée à l'ARN, cette interaction serait impossible. Aussi, la perte d'interaction a été constatée au départ entre les protéines GAL4 DBD-A1 mutante et GAL4 AD-A1 de type sauvage. Lors de l'épissage, il se pourrait que les protéines GST-A1 mutantes seulement aient la capacité d'interagir entre elles.

Suite à l'obtention de ces résultats de modulation de l'épissage alternatif, la perte d'interaction détectée au départ entre les protéines hnRNP A1 mutantes a été réévaluée. Les pertes d'interaction entre les protéines mutantes ont tenté d'être confirmées par une ou plusieurs autres méthodes. Plusieurs méthodes ont été essayées dans cette optique, dont 1) des essais de co-immunoprécipitation entre les protéines de fusion GAL4-A1 produites dans la levure et des protéines GST-A1 produites dans les bactéries, 2) des essais de co-immunoprécipitation entre des protéines de fusion GAL4-DBD-A1 et GAL4-AD-A1 produites dans la levure, 3) des essais de co-immunoprécipitation entre

des protéines GST-A1 et HIS-A1 produites en bactéries, 4) des essais de liaison au glutaraldéhyde (glutaraldéhyde cross-link) entre les protéines HIS-A1, 5) des essais de co-immunoprécipitation entre des protéines hnRNP A1 et myc-hnRNP A1 isolées de lignées cellulaires, 6) des essais de production de protéines hnRNP A1 et GST-A1 de façon polycistronique à partir d'un même vecteur d'expression dans les bactéries et essais de co-immunoprécipitation. Ces différentes tentatives se sont toutes avérées négatives, mais étant donné que les contrôles positifs et négatifs utilisés pour ces différentes techniques n'ont jamais fonctionné, les résultats ne peuvent être ni confirmés, ni infirmés pour le moment.

Les différentes approches de mutagenèse sur la protéine A1 dans le but de déterminer ses domaines protéiques et son mécanisme d'action ont mené à des résultats variables. Certaines mutations ont détruit la capacité d'interaction entre les protéines, et aussi la capacité des protéines mutantes à lier l'ARN. Lorsque ces protéines A1 mutantes capables de lier l'ARN ont été testées dans des essais d'épissage *in vitro*, la modulation observée a été similaire à celle observée en présence de protéine A1 de type sauvage. Il est possible qu'une fois liée à l'ARN, la conformation des protéines mutantes change de manière à permettre à nouveau une interaction entre les protéines, ce qui expliquerait les différents résultats obtenus. Une autre option est de postuler que le modèle est beaucoup plus complexe que ce qui a été élaboré jusqu'à maintenant, et que ces données suggèrent que la liaison entre les protéines A1 n'a pas un grand rôle à jouer dans la modulation de l'épissage alternatif.

CHAPITRE 2

Article 1

Stephen Hutchison, Catherine Le Bel, Marco Blanchette, et Benoit Chabot. (2002)
Distinct Sets of Adjacent Heterogeneous Nuclear Ribonucleoprotein (hnRNP) A1/A2
Binding Sites Control 5' Splice Site Selection in the hnRNP A1 mRNA Precursor. The
Journal of Biological Chemistry 277(33): 29745-29752.

Préambule

L'épissage de l'ARN pré-messager de hnRNP A1/A1^B implique plusieurs niveaux de contrôle afin de préserver le ratio des différentes protéines dans la cellule. Plusieurs régions conservées entre la souris et l'humain sont présentes dans les introns flanquant l'exon alternatif 7B. Certaines de ces régions contiennent des sites de haute affinité pour la liaison de la protéine hnRNP A1 elle-même. Un modèle impliquant ces sites de liaison ainsi que la participation de la protéine A1 elle-même dans le contrôle de la production de son propre ARN messenger a été proposé (Blanchette et Chabot 1999). Dans ce manuscrit, nous décrivons l'analyse de différentes régions contrôlant l'épissage alternatif de l'ARN pré-messager de hnRNP A1 avec l'identification et la caractérisation d'un autre élément (CE1d). La présence de ce nouvel élément dans la séquence conservée CE1 permet de détailler et de renforcer le modèle de « looping-out » proposé auparavant. Nous montrons que hnRNP A1 se lie à cette séquence conservée dans l'ARN pré-messager et que cette liaison est importante pour l'activité de CE1d. Aussi, nous montrons que la protéine hnRNP A2 a la capacité de se substituer à la protéine hnRNP A1 en ce qui a trait à la régulation de l'épissage alternatif d'un substrat modèle, par

liaison des éléments CE1a et CE1d. La présence de multiples sites de liaison pour les protéine hnRNP A1/A2 sur l'ARN pré-messager de hnRNP A1 autour de l'exon alternatif porte à croire que ceux-ci jouent un rôle important dans le contrôle de l'épissage alternatif et que ce mécanisme est finement régulé dans la cellule. Dans cet article, j'ai effectué toutes les expériences concernant la protéine hnRNP A2, son interaction avec l'ARN et son activité dans la modulation de l'épissage alternatif, qui ont mené aux figures 4A, 4B et 4C. Aussi, je travaillais en étroite collaboration avec Stephen Hutchison ce qui m'a permis de travailler sur plusieurs différents aspects de la production des protéines recombinantes (en particulier les protéines GST-A1, GST-A1^B, et GST-A2) qui ont été utilisées dans cet article, ainsi que la vérification des interactions de ces protéines recombinantes avec l'ARN et la vérification de leur activité dans l'épissage alternatif. Malgré le fait que certaines des expériences publiées n'aient pas été effectuées par moi-même, j'ai aussi refait quelques expériences de cet article pour corroborer les résultats obtenus ou pour poursuivre une autre idée. J'ai produit différentes protéines recombinantes (dont celles qui ont mené à la figure 4D et celles qui ont été utilisées dans la figure 2C) et j'ai fait des essais de liaison de ces protéines avec différents transcrits radiomarqués, comme montré dans la figure 2, et des essais d'épissage de certains transcrits, comme dans de la figure 1. Aussi, j'ai effectué plusieurs déplétions d'extraits nucléaires par chromatographie d'affinité (utilisés pour les expériences qui ont mené aux figures 3 et 4), les Western Blot pour vérifier les extraits, et l'épissage *in vitro* avec des protéines recombinantes, comme présenté dans les figures 3 et 4.

Distinct sets of adjacent hnRNP A1/A2 binding sites control 5' splice site selection in the hnRNP A1 pre-mRNA

Stephen Hutchison, Catherine LeBel‡, Marco Blanchette§ and Benoit Chabot¶

Département de Microbiologie et d'Infectiologie, Faculté de Médecine, Université de Sherbrooke, Sherbrooke, Québec, Canada J1H 5N4

RUNNING TITLE: Splicing control by hnRNP A1 and A2

This work was supported by a grant from the Canadian Institute of Health Research (CIHR) (to B. C.).

§Present address: Department of Molecular and Cell Biology, 401 Barker Hall University of California, Berkeley, CA 94720-3204

¶Corresponding author. Mailing address: Département de Microbiologie et d'Infectiologie, Faculté de Médecine, Université de Sherbrooke, Sherbrooke, 3001 12e avenue Nord, Sherbrooke, Québec, Canada J1H 5N4; Phone: (819) 564-5295. FAX (819) 564-5392. E-MAIL b.chabot@courrier.usherb.ca.

ABSTRACT

In the hnRNP A1 pre-mRNA, different regions in the introns flanking alternative exon 7B have been implicated in the production of the A1 and A1B mRNA splice isoforms. Among these, the CE1a and CE4 elements, located downstream of common exon 7 and alternative exon 7B, respectively, are bound by hnRNP A1 to promote skipping of exon 7B *in vivo* and distal 5' splice site selection *in vitro*. Here, we report that CE1a is flanked by an additional high-affinity A1 binding site (CE1d). In a manner similar to CE1a, CE1d affects 5' splice site selection *in vitro*. Consistent with a role for hnRNP A1 in the activity of CE1d, a mutation that abrogates A1 binding abolishes distal 5' splice site activation. Moreover, the ability of CE1d to stimulate distal 5' splice site usage is lost in a HeLa extract depleted of hnRNP A/B proteins, and the addition of recombinant A1 restores the activity of CE1d. Notably, distal 5' splice site selection mediated by A1 binding sites is not compromised in an extract prepared from mouse cells that are severely deficient in hnRNP A1 proteins. In this case, we show that hnRNP A2 compensates for the A1 deficiency. Further studies with the CE4 element reveal that it also consists of two distinct portions (CE4m and CE4p), each one capable of promoting distal 5' splice site use in an hnRNP A1-dependent manner. The presence of multiple A1/A2 binding sites downstream of common exon 7 and alternative exon 7B likely plays an important role in maximizing the activity of hnRNP A1/A2 proteins.

INTRODUCTION

The alternative splicing of mRNA precursors (pre-mRNAs)¹ is a major contributor to the diversity of the mammalian proteome (1-3). The control of splice site selection therefore has profound implications in the production of protein isoforms with different functions. Recent progress in uncovering the molecular strategies that control alternative splicing has led to the identification of many types of sequence elements that influence either positively or negatively the selection of the alternative splice sites. Exonic splicing enhancers are bound by specific members of the SR protein family that can enforce the use of weak 5' and 3' splice sites (reviewed in reference 4). Enhancer elements have also been described in the introns flanking some alternative exons (5-9). Other types of proteins, including members of the hnRNP F/H family of proteins, can bind specifically to intron or exon control elements and hence can contribute to enhancer activity (10-14). Elements that reduce the use of a neighboring splice site are also important in the control of splice site selection. In many cases, the activity of splicing silencers can be mediated by proteins that inhibit specific steps of splice site recognition or spliceosome assembly. A frequent example of this kind of splicing control in mammals involves the mammalian polypyrimidine tract-binding protein (PTB) which binds to some 3' splice sequences and prevents U2AF binding (15-19). Other examples uncovered in mammalian pre-mRNAs include the binding of the SR proteins ASF/SF2 and SRp30c upstream of the branch site in the adenovirus IIIa pre-mRNA and the hnRNP A1 pre-mRNA, respectively (20,21). The mechanisms by which inhibition occurs in these cases appear different since ASF/SF2 prevents U2 snRNP binding in the adenoviral pre-mRNA, while SRp30c does

not prevent the assembly of a U2-containing complex on the downstream 3' splice site (21,22).

Members of the family of core hnRNP A/B proteins have also been identified as factors involved in the modulation of splice site selection. Using model pre-mRNAs carrying competing 5' splice sites, important shifts towards distal 5' splice sites can be obtained by the addition of purified or recombinant hnRNP A1 to a HeLa nuclear extract (23,24). The *Drosophila* hrp48 protein, which is similar to hnRNP A1, is required in collaboration with PSI and the U1 snRNP to repress splicing of the P-element pre-mRNA in somatic tissues (25). HnRNP A1 can elicit exon skipping, but not all pre-mRNAs are responsive to variations in the concentration of hnRNP A1 *in vitro* (26). A1 can also negatively affect the use of 3' splice sites in a variety of exons including the alternative exon of fibroblast growth factor receptor 2 gene (27), the tat and vpr exons of the human immunodeficiency virus (28-30), and the V6 exon of the human CD44 gene (31,32). Consistent with the finding that A1 can recognize specific RNA elements (33), most of the functions that have been attributed to hnRNP A1 in natural pre-mRNAs are based on its ability to interact with specific sequences. In all cases examined to date, the hnRNP A1 splice isoform A1B, the A2 protein and its splicing variant B1 can functionally replace A1 in 5' splice site and 3' splice site selection assays *in vitro* (24,28,29).

We have shown previously that hnRNP A1 can modulate the alternative splicing of its own pre-mRNA through binding to sequences in the introns flanking alternative exon 7B. The 17 nt-long CE1a element downstream of exon 7 and the 24 nt-long CE4 sequence downstream of alternative exon 7B can promote distal 5' splice site selection in an A1-dependent manner (34,35). Both CE1a and CE4 contain the sequence UAGAGU

which closely resembles the “winner” A1 binding site UAGGGU obtained by selection of amplified pools of RNA sequences (33). While these high-affinity A1 binding sites promote strong shifts in 5' splice site utilization *in vitro*, their effect on splicing is not associated with equivalent changes in U1 snRNP binding to the competing 5' splice sites (34). We have proposed that the mechanism by which hnRNP A1 controls 5' splice site selection in this system involves an interaction between bound A1 molecules, an event that would place the proximal 5' splice site in a loop, and would bring in closer proximity the most distant pair of splice sites (35-37). Here, we have uncovered additional elements capable of promoting distal 5' splice site selection *in vitro*. Notably, we find that sequences flanking CE1a and different portions of CE4 can individually interact with hnRNP A1. Our results indicate that control elements promoting distal 5' splice site utilization in the hnRNP A1 pre-mRNA are organized in groups of adjacent A1 binding sites. This organization may facilitate the recruitment of hnRNP A1/A2 proteins and may stimulate or stabilize the proposed change in pre-mRNA conformation.

MATERIALS AND METHODS

Plasmid constructs

The pS1 plasmid was described previously (34). pSCE1 and pSCE1a correspond to the pS2 and pS10 plasmids described in (34). pSCE1z, pSCE1z31, pSCE1z35, pSCE1e, pSCE1d and pSCE1ad were produced by inserting reannealed oligonucleotides corresponding to the mouse sequences of various portions of CE1 at the *Sma I* site of pS1. pC3'-/-, pC5' -/-, pC5' 4/4 and pC5' 1a/1a were described in (35). To generate pC5' 4m/4m, reannealed oligonucleotides corresponding to the mouse CE4m sequence were

inserted into the *Sma I* site downstream of exon 7 and the *EcoR V* site downstream of exon 7B of pC5' -/-. p45 was kindly provided by Faiz Nasim and made by replacing the *EcoO*¹⁰⁹ *I-EcoR I* fragment of pC3' -/- by the *EcoO*¹⁰⁹ *I-EcoR I* fragment from pC5' -/-. p45-WCE1d and p45-M1 through p45-M7 were generated by inserting reannealed oligonucleotides into the *Xma I* site of p45, following Klenow treatment. The reannealed CE1e, CE1d, CE1dM6 and CE1dM7 oligonucleotides were inserted into the *Hinc II* site of pBluescript KS+ to produce pKCE1e, pKCE1d, pKCE1dM6 and pKCE1dM7. pKCE4m and pKCE1a as described previously (35).

Transcription and splicing assays

Splicing substrates were produced from plasmids linearized with *Sca I*, except for p45-M3 which was obtained by cutting with *Ear I*, and transcribed with T3 RNA polymerase (Amersham Pharmacia Biotech) in the presence of cap analog and [α -³²P]UTP (Amersham Pharmacia Biotech). CE1e, CE1d, M6 and M7 RNAs were produced from pKCE1e, pKCE1d, pKCE1dM6 and pKCE1dM7 linearized with *Hind III* and CE1a and CE4m were produced from pKCE1a and pKCE4m linearized with *EcoR I* and transcribed as above. RNA purification was performed as described in (38). HeLa, CB3C7 and CB3C7-20 nuclear extracts were prepared (39) and used in splicing reactions as described previously (34). Creatine kinase was added to HeLa, CB3C7 and CB3C7-20 extracts at a final concentration of 1 unit/12.5 μ l. In all splicing gels, we have used lariat molecules (intermediates and products) to monitor splicing at each 5' splice site. The fact that these molecules usually migrate above the pre-mRNA allows a more precise estimation of splicing efficiency particularly when pre-mRNA degradation obscurs mRNA production. Furthermore, adequate separation between lariat products requires longer runs or the use

of lower percentage acrylamide gels which cause the distal 5' exon and the distal mRNA to run out the gel.

Purification of recombinant proteins

Recombinant GST-A1, A2 and A1B proteins were purified using a glutathione-Sepharose column (Amersham Pharmacia Biotech) following procedure II as described in Fiset and Chabot (40). Bacterial lysis was performed in buffer A (50 mM piperazine-HCl, pH 9.8, 0.5 M NaCl, 1 mM EDTA, 1 mM dithiothreitol, 20 µg/ml bacitracine, 1 mM benzamidine, 0.5 mM phenylmethylsulfonyl fluoride) in the presence of 3 mg/ml lysozyme and 1% Triton X-100. The proteins were eluted from the columns with buffer B (200 mM piperazine-HCl, pH 9.8, 0.5 M NaCl, 1 mM EDTA, 1 mM dithiothreitol, 20 mM reduced glutathione). The purified proteins were dialyzed against buffer D (20 mM Hepes pH 7.9, 100 mM KCl, 20% glycerol, 0.5 mM DTT). The concentration of the recombinant proteins was measured by Bradford (Bio-Rad) using serial dilutions of bovine serum albumin (BSA) as standard.

Gel-shift assays

The RNA samples were heated at 75°C for 5 min in a splicing mix (34) and snapped cooled on ice. The rA1, rA2 and rA1B proteins were then added to the reactions and allowed to incubate on ice for 10 min prior to the addition of heparin (0.74 mg/ml final) and loading dye. The mixtures were run on 6% native acrylamide gels (29:1 acrylamide:bis-acrylamide) in 1X TBE running buffer.

RNA affinity chromatography

The depletion of hnRNP A1/A1BA2/B1 proteins from nuclear extracts was carried out essentially as described in Caputi *et al.* (28). Briefly, fifty nanomoles of synthetic RNA

oligonucleotide corresponding to a sequence within CE1a (UACCUUUAGAGUAGGC) (Dharmacon Research Inc.) were incubated for 1 h at 4 °C, protected from light, in a 100 µl reaction volume containing 100 mM Tris-HCl pH 7.5 and 10 mM sodium periodate. The periodate-treated RNA was coupled to 0.5 ml of agarose adipic acid hydrazide resin following the manufacturer's protocol (Amersham Pharmacia Biotech). The resin was washed twice with 10 ml of storage buffer (20 mM Hepes-KOH pH 7.9, 100 mM KCl, 20% glycerol, 5.7 mM MgCl₂, 1 mM DTT) and kept as a 50% slurry at 4°C. The coupling efficiency, which was typically higher than 95%, was measured by comparing the absorbance at 260 nm of 1% of the input periodate-treated RNA to 10% of the unbound material. 175 µl of HeLa nuclear extract containing 5.7 mM MgCl₂, 0.90 mM ATP, 36 mM phosphocreatine, 3.58 mM DTT and 1.25 unit/ml RNAGuard was incubated with 50 µl of packed beads for 10 min at 30 °C under agitation. The mixture was spun and the supernatant was transferred to a second tube containing 50 µl of the same packed beads. The beads were washed 4 times with 1 ml of 70% buffer D (20 mM Hepes-KOH pH 7.9, 100 mM KCl, 20% glycerol, 1 mM DTT) containing 5 mM MgCl₂. The bound proteins were eluted from the column with 100 µl of loading dye (62.5 mM Tris-HCl pH 6.8, 6 M urea, 10% glycerol, 2% SDS, 0.7 M mercaptoethanol, 0.003 % bromophenol blue).

Western analysis

Equivalent ratios of nuclear extract and proteins eluted from the CE1a RNA column were separated on SDS/10% polyacrylamide gels. These samples were transferred to nitrocellulose filters and probed for the presence of hnRNP A1/A1B/A2/B1 using a rabbit serum containing antibodies against a peptide sequence shared by these proteins (kindly

provided by Telogene Inc, Sherbrooke, Québec, Canada). This serum was used at a dilution of 1:1000 and decorated proteins were revealed using the ECL detection kit (Amersham Pharmacia Biotech) according to the manufacturer's instructions.

RESULTS

Additional sequences within CE1 modulate 5' splice site selection

The 150 nt-long CE1 element is located in the intron downstream of common exon 7 in the hnRNP A1 pre-mRNA. CE1 is required for efficient skipping of alternative exon 7B *in vivo* and stimulates distal 5' splice site use when inserted in between the 5' splice sites of exon 7 and exon 7B in a model pre-mRNA *in vitro* (34). We have reported previously that a small 17 nt segment (CE1a) is responsible for a large portion of the effect of CE1 on 5' splice site selection (34). However, because CE1a does not fully restore the activity of the larger CE1 element, additional sequences within CE1 may also be promoting distal 5' splice site usage. To identify these sequences, we inserted the 3' end portion of CE1 (CE1z, 58 nt; Fig. 1A) between the two 5' splice sites in our model pre-mRNA. Typically, the level of proximal and distal 5' splice site use can be assessed by monitoring the production of lariat molecules following incubation of a labeled pre-mRNA in a nuclear extract and separating the spliced products on a denaturing gel. In comparison to the control S1 pre-mRNA, incubation of SCE1z RNA in a HeLa extract indicated that CE1z stimulated the use of the distal 5' splice site as observed by an increase in lariat products generated by the use of this site (Fig. 1B, compare lane 4 with lane 1). The amplitude of the stimulation was comparable to the effect of CE1a (Fig. 1B, compare lane 4 with lane 3). As observed previously, the full 150 nt CE1 element

promoted a nearly complete loss of proximal 5' splice site use (Fig. 1B, lane 2). To further define the sequence within CE1z that was responsible for the shift, we fragmented CE1z into two smaller slightly overlapping units (CE1z35 and CE1z31). While CE1z31 had a small effect on 5' splice site selection, CE1z35 was much more active in switching splicing to the distal 5' splice site (Fig. 1B, compare lanes 5 and 6 with lane 4). A further dissection of CE1z35 into CE1e (16 nt) and CE1d (19 nt) showed that CE1d, but not CE1e, could shift splicing toward the distal 5' splice site (Fig. 1B, lanes 7 and 8, respectively). These results indicate that the 19 nt-long CE1d element contains sequences that can promote distal 5' splice site selection *in vitro* as efficiently as the previously described CE1a element. When we tested a portion of CE1 that contained both CE1a and CE1d (CE1ad), we observed that the combination of the two elements shifted 5' splice site selection to a level that was comparable to the level obtained with the complete CE1 element (Fig. 1C, compare lane 3 with lane 2). Thus, the activity of CE1 in HeLa extracts apparently results from the combined effects of CE1a and CE1d. (SUGGESTED FIGURE 1 LOCATION)

hnRNP A1 binds to CE1d

Because hnRNP A1 has been implicated in the activity of CE1a, we tested whether hnRNP A1 was also involved in the activity of CE1d. First, we assessed the ability of hnRNP A1 to interact with CE1d by performing a gel mobility shift assay using increasing amounts of a recombinant GST-A1 protein (rA1). The RNA used for this assay is 83 nt-long and contains plasmid sequences followed by the CE1d element. Since CE1e had no effect on 5' splice site selection it was used as a control in place of the

CE1d element. Slower migrating complexes were observed with CE1d but not with CE1e (Fig. 2C, top panel), indicating a specific interaction of rA1 with CE1d.

The observation that recombinant A1 protein can specifically interact with CE1d suggests that this interaction may be required for the CE1d-dependent switch in 5' splice site selection observed *in vitro*. To test this possibility, we produced seven mutations spanning the CE1d element and assessed their effects on 5' splice site selection in a HeLa nuclear extract. The mutations are listed in Figure 2A and they were tested in a slightly shorter version of the S1 pre-mRNA (pre-mRNA 45) for cloning reasons. Inserting CE1d in pre-mRNA 45 (WCE1d) promotes distal 5' splice site utilization in a manner similar to what was observed in the S1 backbone (Fig. 2B, lane 2). The majority of the mutations in CE1d did not compromise the activity of the element (Fig. 2B, compare lanes 3-7 and 9 with lane 2). Mutant M6, however, failed to activate distal 5' splice site selection, yielding a splicing pattern that was similar to the profile obtained in the absence of CE1d (Fig. 2B, compare lane 8 with lane 1). We then tested the small 83 nt-long RNA containing the M6 sequence (M6*) instead of CE1d for binding by hnRNP A1. In comparison to a similar transcript containing the M7 sequence (M7*) which was bound by rA1 (Fig. 2C, bottom panel, lanes 5-8), M6* RNA was not bound significantly by rA1 (lanes 1-4). The correlation between A1 binding and the efficiency of distal 5' splice site activation is consistent with the notion that A1 is involved in the activity of CE1d.

(SUGGESTED FIGURE 2 LOCATION)

The activity of CE1d is mediated by hnRNP A1

To demonstrate that hnRNP A1 mediates the activity of CE1d, we carried out the depletion of hnRNP A1 from a HeLa nuclear extract by affinity chromatography using

RNA molecules carrying a high-affinity binding site for hnRNP A1. The CE1a RNA was covalently linked to agarose adipic acid beads and incubated in the presence of a HeLa nuclear extract. The flow-through fraction was recovered and used as the depleted nuclear extract. Western analysis using a polyclonal antibody that recognizes the core hnRNP A1, A2, B1 and A1B proteins (Fig. 3A, lane 1) indicates that the majority of these hnRNP proteins had been removed from the nuclear extract (lane 5). HnRNP A1, A2, B1 and A1B proteins were found in the fractions bound to successive CE1a columns (Fig. 3A, lanes 6 and 7). A mock-depleted nuclear extract was also prepared by loading a HeLa nuclear extract on a column lacking RNA. The mock-depleted extract contains the core hnRNP A/B proteins (Fig. 3A, lane 2) and no signal was detected in the bound fractions (lanes 3 and 4).

(SUGGESTED FIGURE 3 LOCATION)

The mock- and A/B-depleted extracts were then tested for activity using a CE1a-containing pre-mRNA (SCE1a). We have shown previously that 5' splice site usage on this pre-mRNA can be displaced from the distal to the proximal donor site by the addition of an excess of DNA oligonucleotide carrying A1 binding sites, and that supplementation with rA1 restores efficient distal 5' splice site use (35). Whereas distal 5' splice site use was predominantly observed in the mock-depleted extract (Fig. 3B, lane 1), only the proximal 5' splice site was used in the A/B-depleted extract (lane 2), consistent with a role for hnRNP A/B proteins in the activity of CE1a. SCE1d pre-mRNA was similarly spliced to the proximal 5' splice site in the A/B-depleted extract, suggesting that these proteins may also be required for the activity of CE1d (Fig. 3B, lane 4). The addition of increasing amounts of rA1 to the A/B-depleted extract shifted splicing towards the distal

donor site in a CE1d-dependent manner (Fig. 3C). At the highest concentration tested, the addition of rA1 to the SCE1d pre-mRNA nearly completely abrogated proximal 5' splice site use and activated the distal 5' splice site (Fig. 3C, lane 10). In contrast, at the same concentration of rA1, proximal 5' splice site use remained the predominant choice for the control S1 pre-mRNA (Fig. 3C, lane 6). Thus, the greater sensitivity of a pre-mRNA containing CE1d to the supplementation with rA1 indicates that hnRNP A1 can mediate the activity of CE1d.

HnRNP A2 can also mediate the activity of CE1a and CE1d

To determine whether hnRNP A1 is absolutely necessary for the activity of CE1a and CE1d, we monitored splicing in an extract prepared from a mouse erythroleukaemic cell line previously shown to be severely deficient in hnRNP A1 protein expression. The CB3C7 cell line expresses at least 250-fold lower levels of A1 and A1B mRNAs because one *Hnmpa1* allele has been deleted while the other allele has suffered a retroviral insertion event (41,42). Similar to what was observed in a HeLa extract (Fig. 1B, lanes 1-3), both CE1 and CE1a improved distal 5' splice site utilization in a nuclear extract prepared from CB3C7 cells (Fig. 4A, compare lanes 2 and 3 with lane 1). We also tested splicing in a nuclear extract prepared from a derivative of the CB3C7 cell line which is stably restored for hnRNP A1 expression (CB3C7-20; Fig. 4B, lane 3) (42). The relative levels of distal 5' splice site utilization of the SCE1a pre-mRNA were comparable in the CB3C7-20 and in the A1-compromised CB3C7 extracts (Fig. 4A, lane 5). These results suggest that CB3C7 cells contain a factor(s) that functionally replace hnRNP A1,

allowing CB3C7 extracts to carry out CE1- and CE1a-dependent shifts in 5' splice site selection.

Several observations suggest that the hnRNP A2 protein can functionally replace hnRNP A1 in its ability to promote distal 5' splice site use *in vitro* (24). A2 and its splice variant hnRNP B1 can also replace A1 when repression of a 3' splice site is associated with nearby high-affinity A1 binding sites (28-30). Because CB3C7 cells express hnRNP A2 and B1 (Fig. 4B), these proteins may compensate for the loss of A1 and A1B in CB3C7 cells. We used RNA affinity chromatography to deplete hnRNP A2 and B1 proteins from the CB3C7 nuclear extract (data not shown). The removal of A2/B1 from the CB3C7 nuclear extract was associated with an incapacity to activate the distal 5' splice site upon incubation with the SCE1d pre-mRNA (Fig. 4C, lanes 5 and 13). To confirm that hnRNP A2 could substitute for A1 in the activity mediated by CE1d, we added increasing amounts of recombinant GST-A2 (rA2) or GST-A1 (rA1) to the A2/B1-depleted CB3C7 nuclear extract. At the highest concentrations, the addition of rA1 or rA2 abolished proximal 5' splice site use on the control S1 pre-mRNA (Fig. 4C, lanes 3-4 and 11-12, respectively). The reason for the repression of S1 pre-mRNA splicing by rA1 and rA2 in the mouse extract is unknown. Nevertheless, the addition of rA1 and rA2 had little or no effect on distal 5' splice site use on the S1 pre-mRNA. In contrast, the highest concentrations of rA1 and rA2 tested on the SCE1d pre-mRNA promoted both a reduction in proximal use and an increase in distal splice site selection (Fig. 4C, lanes 7-8, and 15-16, respectively). These results indicate that hnRNP A2 can also mediate the CE1d-dependent activity in 5' splice site selection. Thus, hnRNP A2 is most likely responsible for the lack of a strong splicing defect in the A1-deficient CB3C7 cells.

(SUGGESTED FIGURE 4 LOCATION)

The ability of hnRNP A2 to bind specifically to CE1d was confirmed by performing a gel-shift assay (Fig. 4D). Recombinant hnRNP A2 bound efficiently to CE1d, but only weakly to CE1e (Fig. 4D, top left panel). Moreover, rA2 bound efficiently to the CE1d derivative M7* RNA, but the binding to the M6 sequence was reduced considerably (Fig. 4D, bottom left panel). A binding assay performed with recombinant A1B also indicated specific binding to CE1d (Fig. 5D, top right panel), and stronger binding to M7* relative to M6* RNA (bottom right panel). Thus, CE1d can be bound by A1, A1B or A2. The use of recombinant A1B also allowed a recovery of distal 5' splice site use in a CE1a-dependent manner in hnRNP A/B-depleted extracts (data not shown). Although we have not tested hnRNP B1, the splice variant of A2, we would expect this protein to display a similar binding behavior because it is also depleted from a HeLa extract using a CE1a RNA column. Moreover, several groups have now reported that B1 also displays a function that is similar to hnRNP A1 in splice site selection (24,28).

CE4 also contains several A1 binding sites that influence 5' splice site selection

We reported previously that the 24 nt-long CE4 element downstream of alternative exon 7B contains a high-affinity A1 binding site. CE4 promotes distal 5' splice site selection *in vitro* and exon 7B skipping *in vivo* (35). CE4 contains the sequence UAGAGU (Fig. 5A), which is also present in CE1a, and which was shown to be important for A1 binding and the activity of CE1a (34). Because CE1 contains adjacent A1-bound elements that can individually carry out distal 5' splice site selection, we wondered whether a similar

organization existed in CE4. When we compared the *in vitro* splicing of pre-mRNAs carrying either CE4 or a shortened version lacking the CE4p portion that contains the UAGAGU sequence (see Fig. 5A), we noted a significant reduction in the efficiency of distal 5' splice site use (Fig. 5B, compare lane 2 with lane 4), consistent with the notion that the UAGAGU sequence contributes to the activity of CE4 on 5' splice site selection. On the other hand, the remaining portion of CE4 (CE4m) remained as active as CE1a at promoting distal 5' splice site selection (Fig. 5B, compare lane 4 with lane 3), indicating that CE4m can also promote distal 5' splice site selection *in vitro*.

(SUGGESTED FIGURE 5 LOCATION)

We have noticed earlier that CE4m is not bound efficiently by recombinant hnRNP A1 in a gel-shift assay (35). However, when we repeated the binding assay with recombinant proteins that had been prepared using a high-salt procedure (40), we observed that rA1 could bind to CE4m (Fig. 5C, lanes 6-10). The fact that A1 does not strongly interact with the control CE1e RNA indicates that binding to CE4m is specific (Fig. 5C, lanes 1-5). To assess the contribution of A1 in the activity of CE4m, we tested extracts that had been depleted of A/B proteins by RNA affinity chromatography. The control pre-mRNA C5' -/- was spliced almost exclusively to the proximal 5' splice site in a mock or in the A/B-depleted extract (Fig. 5D, lanes 1 and 2), and the addition of rA1 had no significant effect on 5' splice site selection at the concentrations used (lanes 3 and 4). In contrast, while the C5' 4m/4m pre-mRNA was spliced efficiently to the distal donor site in the mock extract (Fig. 5D, lane 5), distal splicing was abolished in the A/B-depleted extract (lane 6). Moreover, the addition of rA1 to the A/B-depleted extract improved the relative efficiency of distal 5' splice site use to an extent that was

comparable to the level observed in the mock extract (Fig. 5D, compare lane 8 with lane 5). These results indicate that CE4m can also promote distal 5' splice site utilization in an A1-dependent manner. Thus, the strongest activity of CE4 relative to CE1a and CE4m likely reflects the fact that it contains several distinct A1 binding sites.

DISCUSSION

The activity of the previously identified CE1 and CE4 elements can be attributed in each case to adjacent units individually capable of promoting distal 5' splice site selection *in vitro*. CE1 is composed of CE1a and CE1d, each bound by members of the hnRNP A/B family of proteins. Likewise, the activity of CE4 on 5' splice site selection can be separated into CE4m and CE4p, each bound by hnRNP A1 or A2. Recombinant hnRNP A1 and A2 can restore the CE1a-, CE1d- or CE4m-dependent shift in 5' splice site selection in extracts that had previously been depleted of their endogenous hnRNP A/B proteins.

The A1 binding sites in CE1a and CE4p are identical and correspond to UAGAGU, a close match to the optimal A1 binding site UAGGGU sequence obtained by selection from a pool of randomized RNA sequences (33). In contrast, the sequence responsible for the binding of A1 and A2 to CE1d and CE4m is less clear. Although CE4m contains the sequences UAGAUU and UAGACU, we have shown previously that mutating the CE1a UAGAGU into UAGACU compromises A1 binding and abrogates its activity (34). Thus, a pyrimidine at position +5 may be incompatible with efficient A1 binding. CE1d also contains a variety of sequences that matches or resembles the A1 binding consensus UAGRR^{A/U} (R = purines). However, none of the dinucleotide

mutations that hit one of these sites substantially reduced CE1d activity (see Fig. 2). Notably, CE1d and CE4m share the sequence RRGCUAG and ARACU. Moreover, a mutation (M7) that targets these two regions simultaneously affects splicing. This result may indicate the existence of additional non-classical A1/A2 binding sites in CE1d and CE4m.

The existence of several hnRNP A/B binding sites in close proximity could be advantageous for several reasons. First, cooperativity effects due to an increase in the local concentration of A/B proteins may stimulate the apparent K_d of hnRNP A/B proteins. Improving the overall efficiency of A1/A2 binding to these sites could be important when the concentration of A1/A2 proteins becomes limiting (for example, when a cell is actively transcribing many genes). An additional advantage of this organization of A1 binding sites emerges when we consider the model proposing that A1 molecules bound to CE1 and CE4 interact with one another to loop out and repress the internal 5' splice site in a manner that is similar to the effect of duplex-forming RNA sequences (35)¹. The existence of elements (CE1 and CE4) made up of several adjacent A1/A2 binding sites may be similar to having a longer stretch of complementary sequences which improves the rate of duplex formation and offers more stability once the duplex has formed. In the presence of multiple adjacent A1 binding sites, the postulated A1/A1 interactions between CE1 and CE4 could therefore be established more efficiently, and the resulting complex would be more stable. A further gain in stability could be obtained by “cross-strand” interactions, which have been observed in the crystal of a shortened version of A1 (UP1) bound to telomeric DNA sequences (43).

¹ Nasim, F. H., Hutchison, S., Cordeau, M. and Chabot, B., submitted.

This repetitive arrangement of binding sites is a feature that is found in other elements controlling splice site selection. In the *Drosophila* *dsx* enhancer, 6 consecutive 13 nt-long Tra binding sites contribute to the assembly of a highly efficient enhancer complex (44-46). It has been proposed that this organization most efficiently allows the looping of RNA sequence between the enhancer and the 3' splice site region (47). The introns upstream and downstream of alternative exon in the *c-src*, α -actinin, fibroblast growth factor receptor 2 and α -tropomyosin genes also contain several adjacent binding sites for the hnRNP I/PTB protein (reviewed in reference 48).

Several of the functions that have been attributed to hnRNP A/B proteins are associated with the presence of high-affinity binding sites. In addition to a role in the selection of 5' splice sites, many reports have now implicated A/B binding sites in the negative control of 3' splice site choice (28-32). Although the mechanism responsible for this repression has not been examined in all cases, a recent study indicates that an exonic high-affinity A1 binding site can selectively interfere with the binding of some SR proteins to a nearby exon splicing enhancer (49). In this case, the interaction of A1 with the silencer element was associated with the binding of adjacent A1 molecules (49). Thus, although the mechanisms by which exonic and intronic A1 binding sites modulate splice site selection may be different, both situations are apparently mediated by the binding of multiple adjacent A1 molecules.

ACKNOWLEDGEMENTS

We thank Johanne Toutant and Aline Simoneau for the preparation of nuclear extracts, Faiz Nasim for plasmid p45 and members of the Chabot lab for their comments. We thank William Rigby for kindly providing the GST-A2 expression plasmid.

REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001) *Nature* **409**(6822), 860-921.
2. Graveley, B. R. (2001) *Trends Genet* **17**(2), 100-7.
3. Croft, L., Schandorff, S., Clark, F., Burrage, K., Arctander, P., and Mattick, J. S. (2000) *Nat Genet* **24**(4), 340-1
4. Graveley, B. R. (2000) *RNA* **6**(9), 1197-211.
5. Cooper, T. A. (1998) *Mol Cell Biol* **18**(8), 4519-25
6. Huh, G. S., and Hynes, R. O. (1993) *Mol Cell Biol* **13**(9), 5301-14

7. McCullough, A. J., and Berget, S. M. (1997) *Mol Cell Biol* **17**(8), 4562-71
8. Muhlemann, O., Yue, B. G., Petersen-Mahrt, S., and Akusjarvi, G. (2000) *Mol Cell Biol* **20**(7), 2317-25
9. Ryan, K. J., and Cooper, T. A. (1996) *Mol Cell Biol* **16**(8), 4014-23
10. Black, D. L. (1992) *Cell* **69**(5), 795-807
11. Min, H., Chan, R. C., and Black, D. L. (1995) *Genes Dev* **9**(21), 2659-71
12. Min, H., Turck, C. W., Nikolic, J. M., and Black, D. L. (1997) *Genes Dev* **11**(8), 1023-36
13. Caputi, M., and Zahler, A. M. (2002) *EMBO J* **21**(4), 845-855.
14. Caputi, M., and Zahler, A. M. (2001) *J Biol Chem* **276**(47), 43850-9.
15. Southby, J., Gooding, C., and Smith, C. W. (1999) *Mol Cell Biol* **19**(4), 2699-711
16. Lin, C. H., and Patton, J. G. (1995) *RNA* **1**(3), 234-45
17. Singh, R., Valcárcel, J., and Green, M. R. (1995) *Science* **268**(5214), 1173-6
18. Chan, R. C., and Black, D. L. (1997) *Mol Cell Biol* **17**(8), 4667-76
19. Ashiya, M., and Grabowski, P. J. (1997) *RNA* **3**(9), 996-1015
20. Simard, M. J., and Chabot, B. (2002) *Mol Cell Biol* In Press
21. Kanopka, A., Mühlemann, O., and Akusjarvi, G. (1996) *Nature* **381**(6582), 535-8
22. Simard, M. J., and Chabot, B. (2000) *Mol Cell Biol* **20**(19), 7353-62
23. Mayeda, A., and Krainer, A. R. (1992) *Cell* **68**(2), 365-75
24. Mayeda, A., Munroe, S. H., Cáceres, J. F., and Krainer, A. R. (1994) *EMBO J* **13**(22), 5483-95
25. Hammond, L. E., Rudner, D. Z., Kanaar, R., and Rio, D. C. (1997) *Mol Cell Biol* **17**(12), 7260-7
26. Mayeda, A., Helfman, D. M., and Krainer, A. R. (1993) *Mol Cell Biol* **13**(5), 2993-3001
27. Del Gatto-Konczak, F., Olive, M., Gesnel, M. C., and Breathnach, R. (1999) *Mol Cell Biol* **19**(1), 251-60
28. Caputi, M., Mayeda, A., Krainer, A. R., and Zahler, A. M. (1999) *EMBO J* **18**(14), 4060-7
29. Bilodeau, P. S., Domsic, J. K., Mayeda, A., Krainer, A. R., and Stoltzfus, C. M. (2001) *J Virol* **75**(18), 8487-97.
30. Tange, T. O., Damgaard, C. K., Guth, S., Valcarcel, J., and Kjems, J. (2001) *EMBO J* **20**(20), 5748-58.

31. Matter, N., Marx, M., Weg-Remers, S., Ponta, H., Herrlich, P., and Konig, H. (2000) *J Biol Chem* **275**(45), 35353-60.
32. Konig, H., Ponta, H., and Herrlich, P. (1998) *EMBO J* **17**(10), 2904-13
33. Burd, C. G., and Dreyfuss, G. (1994) *EMBO J* **13**(5), 1197-204
34. Chabot, B., Blanchette, M., Lapierre, I., and La Branche, H. (1997) *Mol Cell Biol* **17**(4), 1776-86
35. Blanchette, M., and Chabot, B. (1999) *EMBO J* **18**(7), 1939-52
36. Hutchison, S., Simard, M. J., Nasim, F. H., Cordeau, M., and Chabot, B. (2000) *Recent Res Devel Mol Cell Biol* **1**, 117-135
37. Chabot, B., LeBel, C., Hutchison, S., Nasim, F. H., and Simard, M. J. (2002) in *Progress in Molecular and Subcellular Biology* (Jeanteur, P., ed) Vol. Regulation of Alternative Splicing, In Press, Springer-Verlag GmbH & Co., Heidelberg
38. Chabot, B. (1994) in *RNA processing* (Hames, D., and Higgins, S., eds) Vol. 1, pp. 1-29, Oxford University Press, Oxford
39. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Res* **11**(5), 1475-89
40. Fiset, S., and Chabot, B. (2001) *Nucleic Acids Res* **29**(11), 2268-75.
41. Ben-David, Y., Bani, M. R., Chabot, B., De Koven, A., and Bernstein, A. (1992) *Mol Cell Biol* **12**(10), 4449-55
42. Yang, X., Bani, M. R., Lu, S. J., Rowan, S., Ben-David, Y., and Chabot, B. (1994) *Proc Natl Acad Sci U S A* **91**(15), 6924-8
43. Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R. M. (1999) *Genes Dev* **13**(9), 1102-15
44. Burtis, K. C., and Baker, B. S. (1989) *Cell* **56**(6), 997-1010
45. Nagoshi, R. N., and Baker, B. S. (1990) *Genes Dev* **4**(1), 89-97
46. Tian, M., and Maniatis, T. (1993) *Cell* **74**(1), 105-14
47. Graveley, B. R., Hertel, K. J., and Maniatis, T. (1998) *EMBO J* **17**(22), 6747-56
48. Wagner, E. J., and Garcia-Blanco, M. A. (2001) *Mol Cell Biol* **21**(10), 3281-8
49. Zhu, J., Mayeda, A., and Krainer, A. R. (2001) *Mol Cell* **8**(6), 1351-61.

FOOTNOTES

‡ Recipient of a scholarship from the National Science and Engineering Council of Canada.

¶ B. C. is a Canada Research Chair in Functional Genomics. B. C. is a member of the Sherbrooke RNA/RNP group supported by the FCAR, the Université de Sherbrooke and a CIHR group grant.

¹ The abbreviations used are: pre-mRNAs, mRNA precursors; hnRNP, heterogeneous nuclear ribonucleoprotein; PTB, polypyrimidine tract-binding protein; snRNP, small nuclear ribonucleoprotein particle; PSI, P-element somatic inhibitor; nt, nucleotide; GST, glutathione S-transferase; BSA, bovine serum albumin.

FIGURE LEGENDS

Figure 1. Several elements within CE1 control 5' splice site selection. (A) The structure of the pre-mRNA S1 is represented at the top. The *Sma I* site indicates the position where sequence elements were inserted between the two 5' splice sites. Below is a sequence alignment of the human and mouse CE1 elements. The underlined regions correspond to sequences inserted in pS1 and tested for their ability to influence 5' splice site selection *in vitro*. (B, C) Labeled S1 pre-mRNA derivatives carrying different insertions were incubated in a HeLa nuclear extract for 2 h and splicing products were fractionated on a 11% acrylamide-8 M urea gel. The relative frequencies of distal and proximal splice site utilization can be estimated by comparing the intensity of the bands derived from proximal and distal lariat intermediates and products that migrate above the pre-mRNAs. The position of the pre-mRNAs as well as of the distal and proximal lariat products is indicated. SCE1 and SCE1a have been described previously as S2 and S10, respectively, in Chabot et al. (34).

Figure 2. A mutation within CE1d destroys its activity in splicing and affects A1 binding. (A) Schematic representation of pre-mRNA 45. The arrow indicates the position where mutated CE1d elements were inserted between the two 5' splice sites (*Xma I* site). The sequence of the wild-type and various mutated CE1d elements is shown below. Uppercase letters correspond to the inserted sequences, mutations are indicated in bold/underlined letters. (B) Splicing reactions were set up in HeLa extracts and products were separated on a 12% acrylamide-8 M urea gel. The position of the distal and proximal lariat molecules is shown. Note that pre-mRNA/mRNA M3 is shorter than the

other mutants because *Ear I* instead of *Sca I* was used to linearize the corresponding plasmid. (C) Binding of hnRNP A1 to CE1d and mutated derivatives. Radiolabeled RNAs containing the CE1e and CE1d (top panel) and the M6 and M7 mutated elements (bottom panel) were incubated with increasing amounts (0, 0.125, 0.25 and 0.5 μ g) of GST-A1 (rA1). Mixtures were fractionated in non-denaturing gels. The position of the free RNA and the RNA/protein complexes is shown.

Figure 3. HnRNP A1 mediates the activity of CE1d. (A) Depletion of hnRNP A1/A1B/A2/B1 proteins from a HeLa nuclear extract by RNA affinity chromatography. A HeLa extract was loaded onto an agarose adipic acid column covalently linked to a portion of CE1a. A control mock depletion was run in parallel. Western analysis of the CE1a- and mock-depleted HeLa nuclear extracts was carried out with an antiserum raised against a peptide shared by hnRNP A1, A1B, A2 and B1 proteins. Δ corresponds to a nuclear extract that has been depleted of hnRNP A/B proteins by chromatography on a CE1a column, while M corresponds to a mock-depleted extract. Bd1 and Bd2 correspond to fractions eluted from two successive CE1a or mock columns. The input sample consists of an equivalent fraction of the initial HeLa extract. The lower panel represents an overexposure of the upper panel. We estimate that less than 5% of A1/A1B/A2/B1 proteins remains in the A/B-depleted extract. (B) Splicing in depleted extracts. Radiolabeled pre-mRNAs containing either CE1a or CE1d were spliced in the mock- (M) and A/B-depleted (Δ) extracts. (C) rA1 add-back experiment. Distal 5' splice site selection was tested with S1 and SCE1d pre-mRNAs with increasing amounts of rA1 (0, 0.5, 1 and 2 μ g) added to a A/B-depleted HeLa nuclear extract. Both S1 and SCE1d were

also spliced in the mock-depleted nuclear extract (lanes 1 and 2). Splicing products in panels B and C were fractionated on a 11% acrylamide-8 M urea gel. The position of pre-mRNAs, proximal and distal lariat products is shown.

Figure 4. The activity of CE1d can also be mediated by hnRNP A2. (A) Splicing of S1, SCE1a and SCE1 pre-mRNAs in CB3C7 and CB3C7-20 nuclear extracts. CB3C7 are mouse erythroleukaemic cells that are deficient in A1 expression. A1 expression was stably restored in CB3C7-20 cells. (B) Western analysis of A1/A2 expression in CB3C7 and CB3C7-20 cells. The asterisk (*) represents an uncharacterized band detected only in the mouse erythroleukaemic nuclear extracts. (C) Splicing of the S1 and SCE1d pre-mRNAs in a CB3C7 nuclear extract depleted by chromatography on a CE1a column, and supplementation with increasing amounts of rA1 or rA2. To each set, 0, 1, 2 and 3 µg of the recombinant protein was added. (D) Radiolabeled RNAs containing the CE1e, CE1d, M6 and M7 sequences were tested for their ability to interact with rA2 and rA1B. Each set consisted of 0, 0.125, 0.25, 0.5 and 0.75 µg of rA1B or rA2. Mixtures were fractionated on native polyacrylamide gels. The position of the RNA/protein complexes and free RNA is indicated.

Figure 5. CE4m also displays A1-dependent 5' splice site modulating activity. (A) Schematic representation of the C5' -/- transcript. The sequence of CE1a, CE4 and CE4m are shown and each one was inserted at both positions in C5' -/- indicated by arrows. The high-affinity A1 binding site UAGAGU is indicated by bold and underlined letters. (B) Various pre-mRNAs were incubated in a HeLa extract and splicing products were

fractionated as described earlier. (C) A radiolabeled RNA containing the CE4m sequence was compared with CE1e RNA for its interaction with rA1 in a gel-shift assay. Each set was tested with 0, 0.125, 0.25, 0.50 and 0.75 μ g of recombinant protein. (D) Splicing reactions were performed in a mock- (M) or an A/B-depleted (Δ) HeLa nuclear extract using the C5' -/- or the C5' 4m/4m pre-mRNA. Increasing amounts of rA1 (0.5 and 1 μ g) were also added to the Δ extract. Splicing products were separated on an 11% acrylamide-8 M urea gel. Note that proximal lariat molecules from C5' 4m/4m, C5' 1a/1a and C5' 4/4 migrate above the pre-mRNA because their loop size is larger than the lariat derived from the C5' -/- transcript.

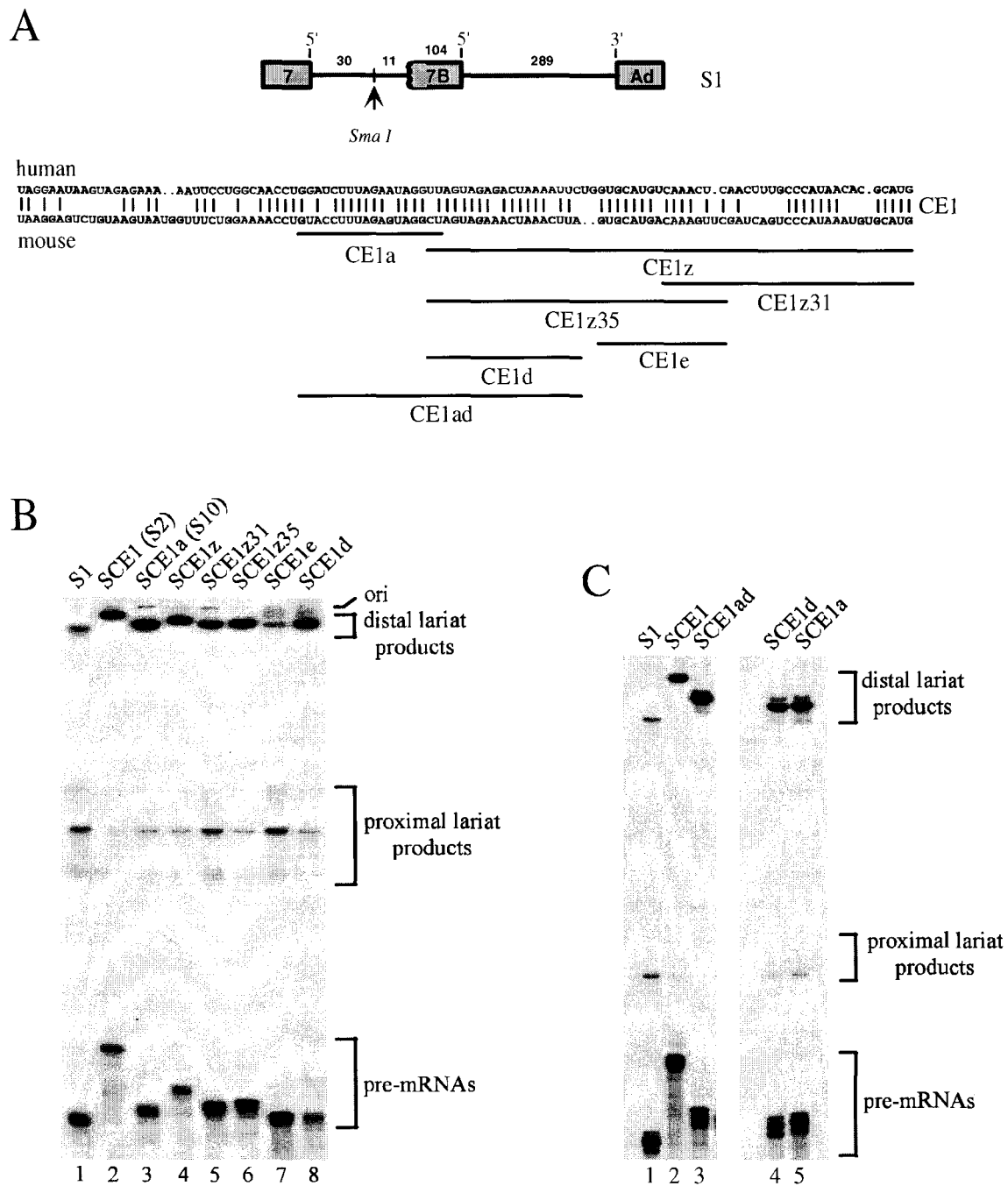


Figure1 A-B-C. Hutchison, S., LeBel, C., Blanchette, M., and Chabot, B.

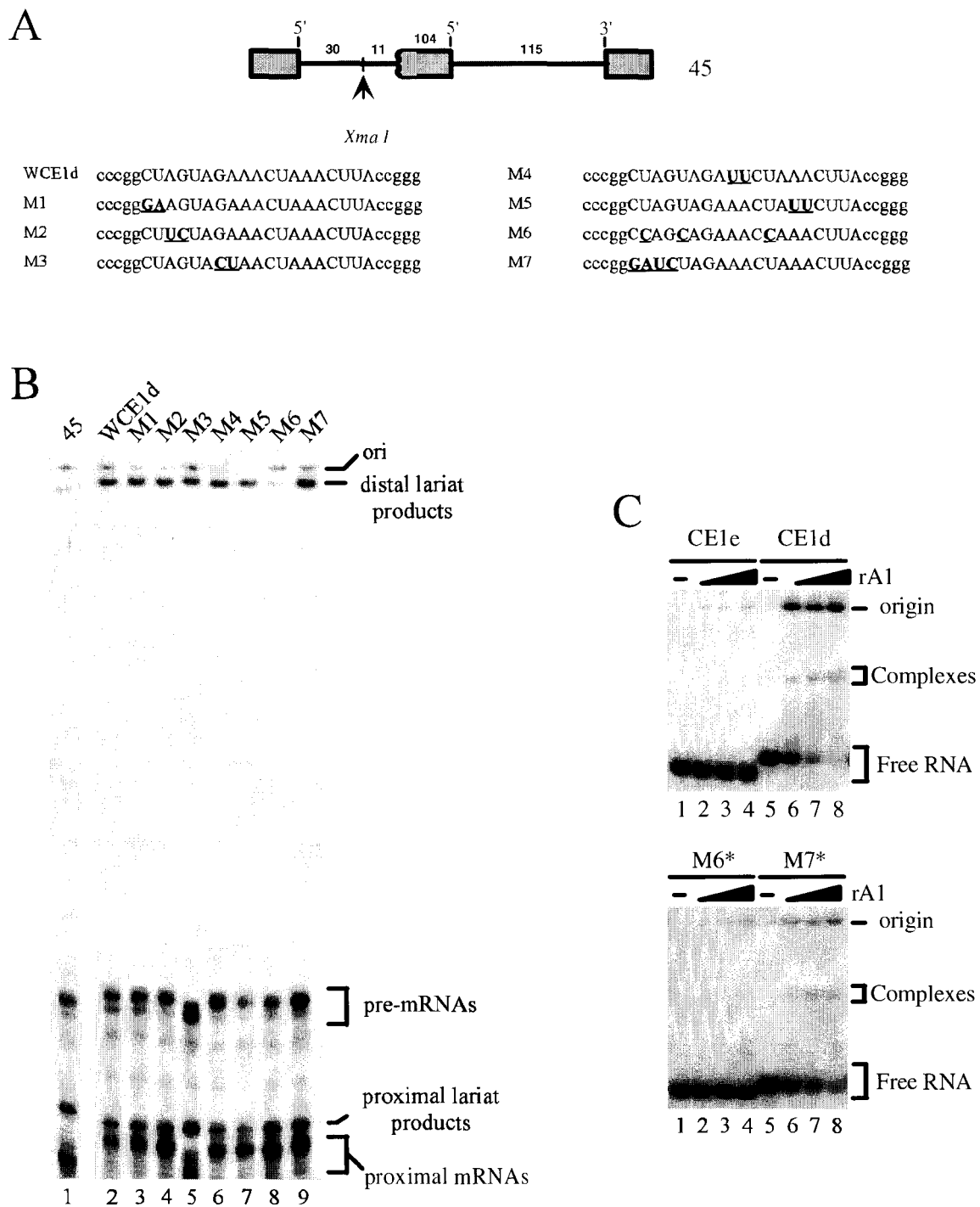


Figure 2 A-B-C. Hutchison, S., LeBel, C., Blanchette, M., and Chabot, B.

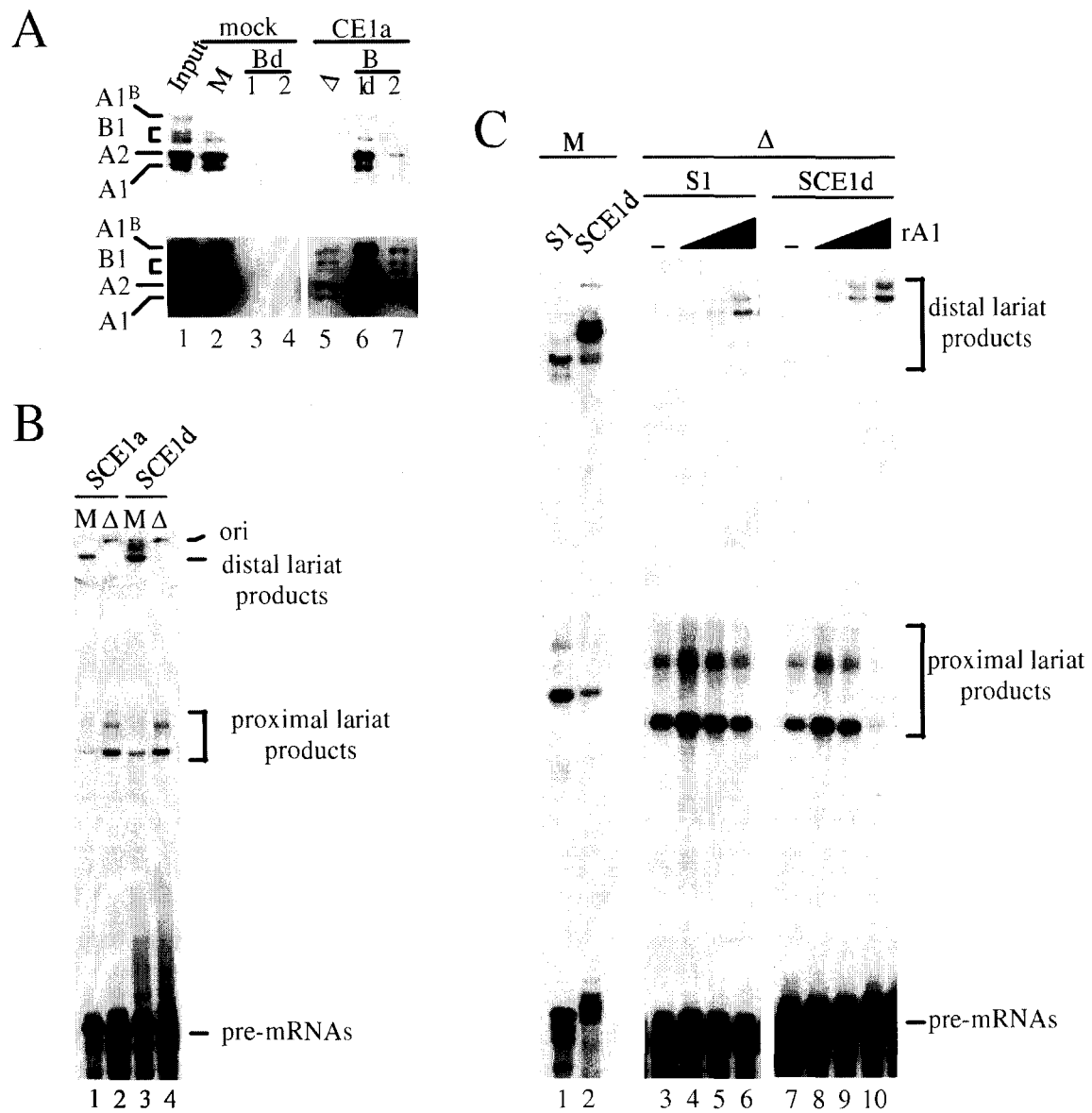


Figure 3 A-B-C. Hutchison, S., LeBel, C., Blanchette, M., and Chabot, B.

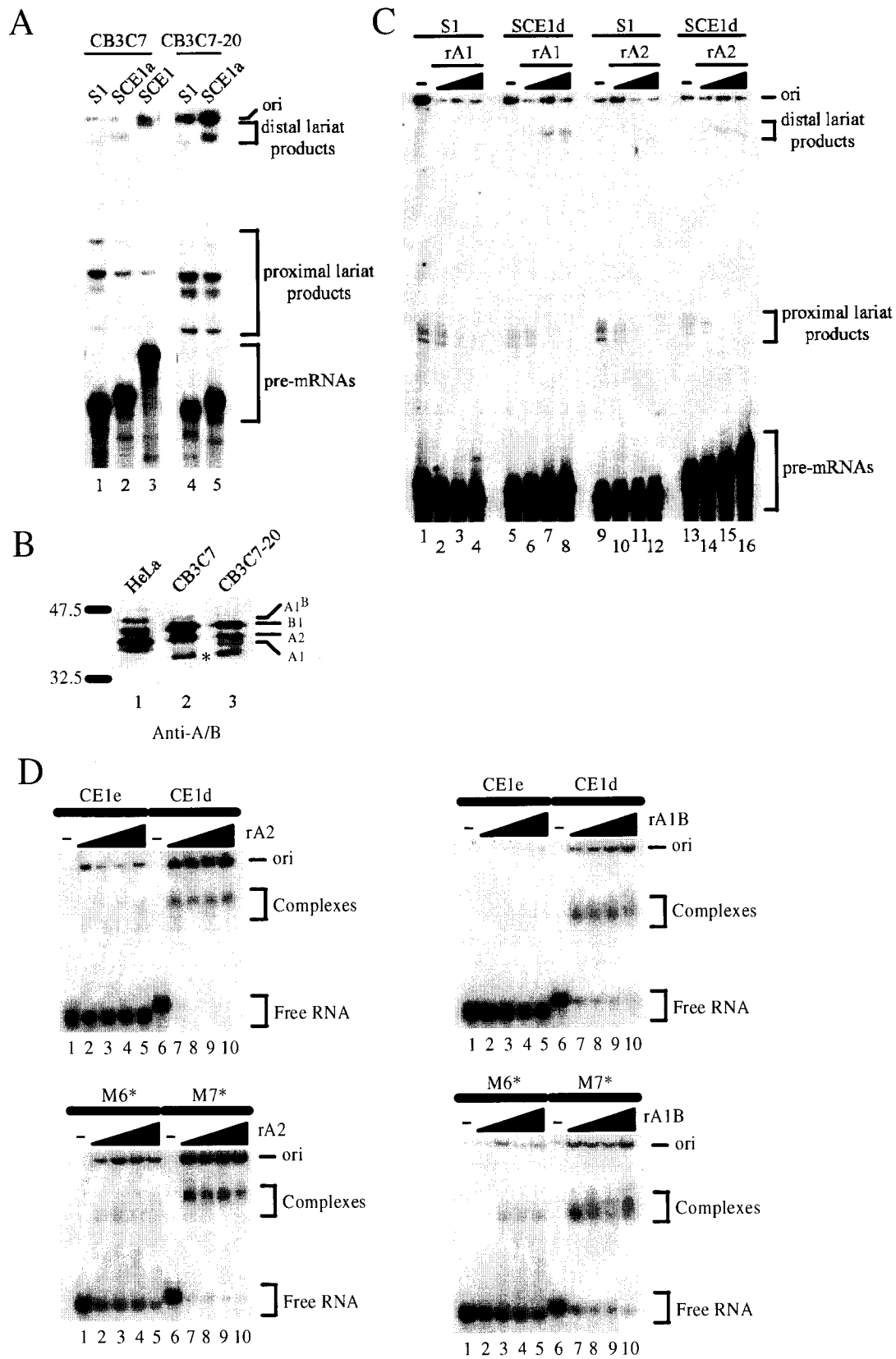


Figure 4 A-B-C-D. Hutchison, S., LeBel, C., Blanchette, M., and Chabot, B.

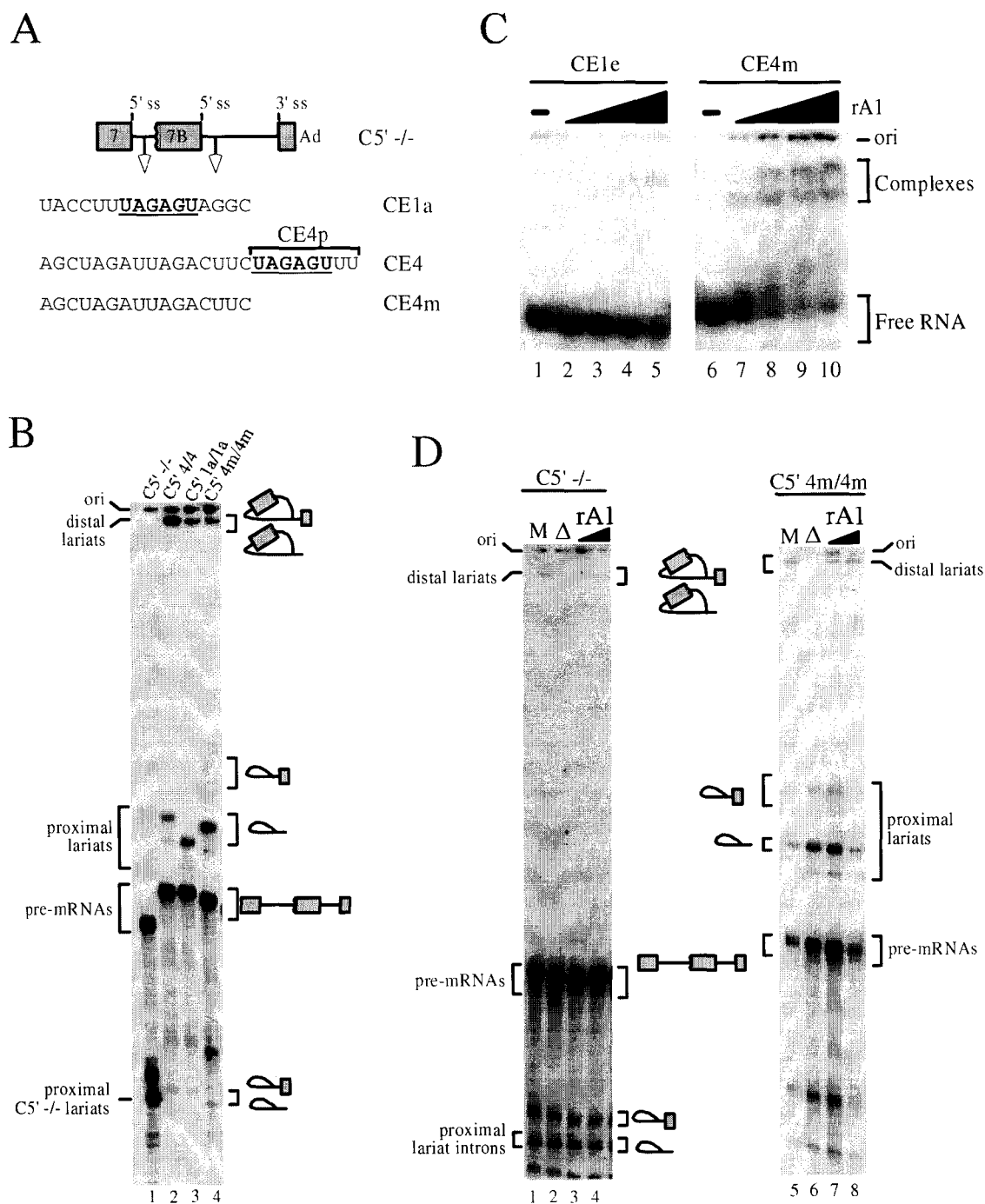


Figure 5 A-B-C-D. Hutchison, S., LeBel, C., Blanchette, M., and Chabot, B.

Article 2

Véronique Bériault, Jean-François Clément, Kathy Lévesque, Catherine Le Bel, Xiao Yong, Benoit Chabot, Eric A. Cohen, Alan W. Cochrane, William F. C. Rigby, et Andrew J. Mouland. (2004) A Late Role for the Association of hnRNP A2 with the HIV-1 hnRNP A2 Response Elements in Genomic RNA, Gag, and Vpr Localization. The Journal of Biological Chemistry 279 (42): 44141-44153

Préambule

Cet article est inséré dans le chapitre 2 et est un complément à l'article mentionné auparavant. Ce manuscrit permet aussi d'illustrer la collaboration entre différents laboratoires appartenant au Ribo-Club. La protéine hnRNP A2 possède différentes fonctions dans la cellule et est impliquée dans la modulation de la sélection des sites d'épissage dans des essais *in vitro*, mais elle est plus particulièrement impliquée au niveau de la transcription et du transport de l'ARN. Cet article montre un rôle de la protéine hnRNP A2 dans le transport et la localisation de différents ARN messagers, contenant des sites de liaison pour hnRNP A2. Dans cet article, le travail que j'ai accompli fut la préparation de différents transcrits, basés sur les transcrits modèles utilisés dans le laboratoire pour l'étude de l'épissage alternatif avec la protéine hnRNP A1 (Blanchette et Chabot 1999; Hutchison et al. 2002; Chabot et al. 2003), mais dans lesquels les sites de liaison de haute affinité pour hnRNP A2 ont été insérés (A2RE). Ensuite, j'ai effectué l'épissage alternatif *in vitro* de ces transcrits. Les réactions d'épissage alternatif utilisaient des extraits nucléaires normaux ou des extraits nucléaires dépourvus des protéines hnRNP A1/A1^B/A2/B1. J'ai aussi purifié les différentes

protéines recombinantes utilisées lors des essais d'épissage dans des extraits nucléaires dépourvus, et j'ai effectué les comparaisons entre les transcrits modèles, afin de mieux caractériser l'action des protéines hnRNP A1 et A2 au niveau de la régulation de l'épissage alternatif. Cet article se trouve intégré dans la section annexe.

CHAPITRE 3

Michel Larrivée, Catherine Le Bel, et Raymund J. Wellinger. (2004) The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. *Genes Dev.* 18: 1391-1396.

Préambule

Ce manuscrit rapporte pour la première fois dans la littérature, la présence d'extensions G-riche simple-brin présentes aux télomères de la levure *Saccharomyces cerevisiae*, en phase G1 du cycle cellulaire. La démonstration de la présence de ces extensions est faite avec l'utilisation de deux techniques totalement différentes; la plus connue est l'hybridation « in-gel » et la seconde, que j'ai développée et mise au point, est une extension d'amorce modifiée. Cette méthode permet d'obtenir une taille précise pour la longueur maximale des extensions télomériques présentes sur une population de télomères. De plus, nos résultats démontrent hors de tout doute ce qui était seulement présumé depuis plusieurs années, c'est-à-dire l'existence d'extensions télomériques constitutives sur la vaste majorité des télomères de levure en dehors de la phase S. Nos résultats indiquent également que le complexe protéique MRX est essentiel pour la formation des extensions télomériques normales chez la levure; en son absence, les extensions simple-brin présentes aux télomères sont plus courtes (8 bases vs 12-14 bases), mais semblent toujours être générées durant la phase S. Ces résultats suggèrent que la présence et la formation adéquate des extensions télomériques simple-brin sont critiques pour la cellule, et que ces extensions sont générées d'une autre façon chez les cellules où le complexe MRX est absent. Ceci implique qu'il y aurait des exonucléases

télomériques redondantes au complexe MRX. M. Larrivée et moi avons chacun effectué environ 50% des expériences rapportées dans cet article. Plus spécifiquement, j'ai effectué toutes les expériences relatives aux figures 3 et 4A, et j'ai aussi effectué plusieurs hybridations « in gel » avec de l'ADN génomique de différentes souches de levures, comme montré dans la figure 1.

The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX-complex.

Michel Larrivée, Catherine LeBel and Raymund J. Wellinger *

Département de Microbiologie et Infectiologie
Groupe ARN/RNA Group, Faculté de Médecine
Université de Sherbrooke
3001 12^e Avenue Nord
Sherbrooke, J1H 5N4
Canada.

Running Title: MRX is essential for G-tails in yeast

Keywords: Telomeres, G-tails, MRX-complex, Yeast.

*Corresponding author. Phone: (819) 564-2514. Fax: (819) 564-5392. E-mail:

Raymund.Wellinger@Usherbrooke.ca

ABSTRACT

The precise DNA arrangement at chromosomal ends and the proteins involved in its maintenance are of crucial importance for genome stability. For the yeast *Saccharomyces cerevisiae*, this constitutive DNA configuration has remained unknown. We demonstrate here that G-tails of 12-14 bases are present outside of S-phase on normal yeast telomeres. Furthermore, the Mre11p protein is essential for the proper establishment of this constitutive end-structure. However, the timing of extended G-tails occurring during S-phase is not affected in strains lacking Mre11p. Thus, G-tails are present on yeast chromosomes throughout the cell cycle and the MRX-complex is required for their normal establishment.

INTRODUCTION

The physical ends of eukaryotic chromosomes, the telomeres, have a very conserved structure and are essential for genome stability (reviewed in Blackburn 2001; Chakhparonian and Wellinger 2003). Short direct DNA repeats constitute the underlying telomeric DNA and the strand running 5' to 3' toward the end of the chromosomes is usually rich in guanines (the G-rich strand). Lagging-strand synthesis always occurs on this G-rich strand and will leave a short gap at the 5' end of the newly synthesized C-rich strand. This gap can not be filled in by repair and a 3' G-rich overhang, called G-tail, remains. On the other end, leading-strand synthesis is thought to produce a blunt extremity. However, studies of the terminal DNA arrangement in a variety of organisms suggest that a G-tail is a conserved motif for all telomeres (Chakhparonian and Wellinger 2003). Thus, the question arises as to how the blunt-ended DNA-ends generated by leading-strand synthesis are converted into ends with a G-tail.

Studies in the yeast *Saccharomyces cerevisiae* have shown that its telomeres acquire detectable G-tails late in S-phase, after conventional replication (Wellinger et al. 1993a; Wellinger et al. 1993b). Moreover, at least on the ends of a linear plasmid, G-tails occur on both, leading- and lagging-strand ends (Wellinger et al. 1996). Surprisingly, these S-phase specific G-tails can also be detected in cells lacking telomerase, the main activity responsible for replicating telomeric G-strands (Dionne and Wellinger 1996). Collectively, these results suggest that the blunt end left after completion of leading-strand synthesis is processed into an end with a G-tail, presumably by nuclease/helicase activities (Wellinger et al. 1996). Analyses of the requirements to establish a normal

telomeric DNA end-structure are hampered by the fact that for wild-type (wt) yeast cells, the precise DNA arrangement outside S-phase is unknown.

Recent studies on the Mre11p/Rad50p/Xrs2p (MRX) proteins, an evolutionarily conserved complex involved in a number of processes in mitosis and meiosis, revealed that this complex may play a key role in telomere length maintenance in humans, plants and yeasts (reviewed in Haber 1998; D'Amours and Jackson 2002). Yeast cells harbouring a deletion of any one of these genes are viable, but display shortened telomeric repeat tracts (Kironmai and Muniyappa 1997; Boulton and Jackson 1998). The Mre11p protein alone, or in association with other proteins, displays various nucleolytic activities (D'Amours and Jackson 2002), and it has been demonstrated that in a *de novo* telomere formation assay, the MRX-complex is needed for the generation of telomeric G-strand DNA and the loading of G-tail binding proteins (Diede and Gottschling 2001). In addition, the MRX-complex in *Schizosaccharomyces pombe* might process telomeric DNA ends in absence of a DNA double-stranded binding protein Taz1p (Tomita et al. 2003).

In order to gain insights into the requirements to establish a normal DNA-end at telomeres, we developed a stringently controlled hybridization assay as well as a primer-extension assay. The results demonstrate that most of normal yeast telomeres end in a G-tail of about 12-14 bases. Moreover, we show that in *mre11Δ*-strains, this constitutive DNA end-structure is compromised; specifically, G-tails in such mutant strains are shorter. In addition, while the dynamics of the cell cycle-dependent increase of G-tail signals at the end of S-phase occurs in both wt and mutant strains in an indistinguishable fashion, signals for G-tails at the end of S-phase were also weaker in the *mre11Δ*-strain.

The data thus show that short 10-15 bases G-tails are the proper ends of yeast chromosomes and in *mre11Δ* cells, there is a deficiency in forming those G-tails, even though G-tail presence is not completely abolished. Therefore, the MRX-complex does play a direct role in telomeric end-structure processing in normal yeast cells and could play a similar role at mammalian telomeres.

RESULTS AND DISCUSSION

Telomeric end-structure is affected in the absence of MRX-complex. There is accumulating evidence that the three interacting proteins Mre11p, Rad50p and Xrs2p (NBS1), forming the MRX- or MRE11-complex, play a role in telomere maintenance in yeasts, plants and mammals. If the MRX-complex were involved in the processing of the blunt ends left after completion of leading-strand synthesis at the telomeres, one would expect to detect shorter G-tails (either transiently or constitutively) on yeast telomeres in strains lacking any of the components of the complex. The presence of G-tails can be assessed using a non-denaturing in-gel hybridization technique that we developed previously (Dionne and Wellinger 1996). Such analyses on DNA derived from asynchronously grown wild-type (wt) cell cultures yielded very faint signals for chromosomal terminal restriction fragments (TRFs). It remained unclear, whether these signals were due to constitutive short G-tails or whether they were due to a fraction of cells being in late S-phase, when G-tails are easily detectable on yeast telomeres (Wellinger et al. 1993b; Dionne and Wellinger 1996). However, such signals on telomeres of a high-copy linear plasmid of 7.5 kb, called YLpFAT10, are concentrated in a small area of the gel and become readily visible, even on DNA derived from G1-arrested cultures (fig. 1A, middle; fig. 1B). FACS analysis and counting of unbudded vs. budded cells of these cultures established that the vast majority of cells ($\geq 95\%$) were indeed in G1 (data not shown, see Supplementary fig. 1). More importantly, when genomic DNA isolated from *mre11* Δ strains was analyzed in the same fashion, a significant reduction of the signals for the G-tails is observed (fig. 1A, middle panel, compare lanes marked *MRE11* with those marked *mre11* Δ ; fig. 1B). Treatment of the

DNA with exonuclease I prior to the end-structure analyses (lanes ExoI+ in fig. 1A) indicates that the detected signal indeed corresponds to terminal single-stranded G-strands. Furthermore, the same gel shown in fig. 1A was also hybridized to a G-rich probe and no telomeric C-rich signals were detected (data not shown). Quantification of the signals for G-tails confirmed that they are about 1.3-2 fold lower on DNA derived from *mre11Δ*-cells as compared to wt cells (fig. 1B). Similar results were obtained for a *rad50Δ* strain, with strains of different genetic backgrounds and when chromosomal telomeres were analyzed (fig. 1B, supplementary fig.2, and data not shown). Collectively, these results demonstrate that in G1, G-tails can be detected on telomeres of yeast cells, and that the signal for G-tails in *mre11Δ*-cells is reduced when compared to wt cells.

We next investigated the length of the telomeric G-tails for cells in G1. As a means of comparison, constructs containing a stretch of nucleotides corresponding to telomeric repeats were used. The plasmid with the longest tract (GT22) contains 22 nucleotides that are perfectly complementary to the CA-probe used for the in-gel hybridization technique. In addition, derivatives of GT22 were obtained (GT19 to GT10). These constructs harbour stepwise shortened tracts of telomeric sequences, each being 3 nucleotides shorter than the previous one. The different plasmids generated were double-digested and heat denatured prior to loading onto a gel (lanes GT10 to GT22 in fig. 2A). Note that a fragment migrating to an apparent length of 1.1 kb of all plasmids carries a sequence complementary to T7 and M13 probes, and that the G-rich telomeric repeat tract of indicated size resides on slightly larger fragment. On the same gel, genomic DNA derived from G1-arrested wt-cells was analyzed for G-tails (lanes WT in fig. 2A). The gel was first hybridized to an M13 probe to show about equal loading of the GTxx plasmids

(fig. 2A, top left, open arrowhead). Next, the gel was hybridized to the 22-mer CA probe, revealing the fragment containing the G-rich strand of the different GT-plasmids (closed arrowheads in fig. 2A), as well as G-tails on YLpFAT10 (* in fig. 2A). An exposure was obtained after each washing step and the relative signals for all bands were quantified. By comparing the loss of signals occurring at each individual washing temperature for the known lengths of GT-tracts on the plasmid DNA to the loss of signals on YLpFAT10 telomeres, an estimation of the lengths of G-tails on YLpFAT10 could be derived (fig. 2B). Clearly, at lower washing temperatures (20°C-30°C), the signals for the telomeres on YLpFAT10 are reduced to a greater extent than the signals for even the shortest model target plasmid, GT13 (red vs black curve in fig. 2B). If the overhangs were completely complementary to the 22-mer CA probe, these results suggest that they are shorter than 13 nt. However, given that telomeric overhangs are of a variable sequence composition, it is likely that a considerable fraction of telomeres contains a number of mismatches with respect to the probe. We compared all theoretical possibilities of 16 nt overhangs to the sequence of the probe and found that all can yield at least 13 nt matches to it (data not shown). Thus, this experiment suggests that G-tails on telomeres of wt cells are shorter than 16 bases and only a minority could bear G-tails longer than that.

The above provided a rough estimate for the lengths of G-tails on a population of telomeres in wt cells, but could not yield a precise determination of G-tail lengths on individual telomeres. We therefore developed a modified primer-extension method, in which poly-A tailed purified TRFs were annealed to an end-labelled poly-T oligonucleotides ending with a C or an A at their 3'-end (fig. 3, top). The annealing products were then extended using T4 DNA polymerase in conditions that minimized

strand-displacement at the double-stranded to single-stranded transition point. In order to verify the reliability of the method, conditions were established using a model oligonucleotide substrate with a sequence composition similar to the one expected on the TRFs (DUP-16, fig. 3 left). TRFs purified from the various strains were first analyzed for telomeric DNA end-structure using non-denaturing in-gel hybridization to make sure that no alterations occurred during DNA manipulations (data not shown). Optimized conditions were then applied to such purified TRFs derived from wt and *mre11Δ*-cells (fig. 3). For wt cells, extension products up to 12 to 14 nucleotides are readily detectable. Remarkably, products of 9 to 12 nucleotides were more abundant than shorter products. In contrast, short products (+ 2 to + 8) were much more abundant than the longer products when DNA derived from *mre11Δ*-cells was used (fig. 3). However, we note that even with DNA derived from these mutant cells, weak signals for G-tails of the same length as those seen with DNA derived from wt cells could be detected. Quantification of the signals revealed that with DNA derived from wt cells, about 3-4 fold more signal is in the range of +8 and longer as compared to the *mre11Δ*-cells (data not shown, see fig. 3). Therefore, while all strains can have G-tails of 12 to 14 bases, the fraction of telomeres ending with a G-tail of 8 nt and longer is significantly reduced in *mre11Δ*-cells. At least qualitatively, these results are consistent with the results obtained using hybridization techniques (figs. 1, 2) and confirm that a large fraction of the constitutive G-tails observed on telomeres in G1 is shorter in *mre11Δ*-cells, when compared to wt cells. Taken together, the data show that the constitutive chromosomal DNA end-structure for wt yeast cells encompasses overhangs of the G-rich strand that vary in length with most

of them being shorter than 14 nucleotides. Moreover, the constitutive G-tails on telomeres of *mre11* Δ -cells clearly are shorter, most of them being shorter than 8 bases.

The majority of yeast telomeres carry constitutive overhangs. Next, we wished to examine what fraction of telomeres does carry such 3'-overhangs and whether some telomeres carried a blunt end or a short 5'-overhang. Genomic DNA with a native telomeric end-structure was treated with Exonuclease III. This strand-specific exonuclease degrades the strands with the 3'-end on dsDNA that is blunt or possesses a 5'-overhang. As positive controls, we used genomic DNA on which telomeres were blunt-ended by ExoI treatment as in fig. 1 prior to ExoIII treatment. On this DNA, the generation of 5'-overhangs is detectable by a GT-probe after non-denaturing in-gel hybridization (fig. 4A). However, native ends are by en large resistant to ExoIII treatment (fig. 4A). Quantification of the gel revealed that the signals for blunted ends were 5-10 fold increased when compared to those obtained with native telomeres. In a second and independent approach, we used YLpFAT10 telomeres to determine the amount of ends harboring G-tails. Since YLpFAT10 as well as the pGTxx plasmid series each contained one sequence element complementary to the T7-probe, the estimated number of telomeres ending with G-tails could be derived by comparing signals obtained on pGTxx plasmids with those on YLpFAT10. DNA in gels such as shown in fig. 2A was denatured, hybridized to a T7 probe and the signals for the GT22, GT19 plasmids and YLpFAT10 were quantified. This yielded an approximate ratio of YLpFAT10 molecules with respect to the control pGTxx DNAs on the gels. In parallel, the same DNA was analyzed by non-denaturing in-gel hybridization using the CA-telomeric probe, washed at 4°C and signal intensities for plasmids GT22, GT19 and YLpFAT10 were again

quantified. These latter signal intensities were corrected for the fact that each YLpFAT10 molecule contains two telomeres whereas the pGTxx plasmids only contained one target sequence for the CA-probe. The obtained values were then adjusted to the molecular ratio of the molecules. Assuming equal chances of hybridization to each of the targets, the obtained values indicate the fraction of YLpFAT10 telomeres hybridizing to the probe (fig. 4B). Although there is some variability, the data corroborate the results obtained with the ExoIII experiments in that at least 80% of the telomeres on YLpFAT10 hybridized to the probe, when the gels were washed at 4°C. Taken together, these data indicate that the vast majority and most likely all native telomeres have a 3'-overhang.

The dynamics of the cell cycle-dependent changes in G-tail detection is not affected in mre11Δ-cells. Since we observed that *mre11Δ*-cells display shorter G-tails than wt cells, we were interested to determine whether the S-phase dependent increase in G-tail signals observed in wt cells was decreased or abolished in the mutants. *MRE11*- and *mre11Δ*-cells were synchronized and DNA analyzed at different time points by in-gel hybridization (suppl. figs. 1 and 2; data not shown). A quantification of the relative signals shows a significant reduction for telomeric G-tail signals in α -factor and *cdc7*-arrested *mre11*-cells when compared to wt cells (fig. 5). For the cells that were released, the relative signals for G-tails increase for both strains in S-phase (fig. 5). From these data it is clear that *mre11Δ*-cells display a cell cycle-dependent change in the G-tail signals as wt cells do. Throughout the experiment, a significant difference between G-tail signal intensity for wt and *mre11Δ* strains was observed, suggesting that not only G1-specific G-tails, but also the S-phase specific G-tails may be shorter in the mutant strain. However, due to the techniques used, such comparisons between cultures are challenging

and it remains uncertain whether this constant difference is real. Nevertheless, these results are consistent with previous data, which suggested that nucleolytic processing of a DNA double-strand break generated at the *MAT* locus during mating-type switching is inefficient in strains harbouring deletions of the *RAD50* or *XRS2* genes (Ivanov et al. 1994). It is therefore tempting to speculate that the lower levels of the cell cycle-dependent G-tails observed in our synchrony experiments with *mre11Δ*-cells are due to inefficient processing. Importantly though, the data do show that in the mutant strain, S-phase dependent long G-tails are generated, suggesting that it is a processing event that probably occurs after DNA replication and which should establish a normal end-structure for the rest of the cell cycle, that is affected in these cells.

There is previous evidence that the MRX-complex could be involved in a chromosome-end processing event. For example, with respect to G-strand replication, the MRX-complex is in the same epistasis group as telomerase components (Nugent et al. 1998). Furthermore, in a telomere formation assay that is based on HO-cutting and exposing a short telomeric repeat tract, the MRX-complex was shown to be needed for Cdc13p-binding, and hence G-tail formation. In *mre11Δ*-cells, telomere formation was delayed, but not completely abolished (Diede and Gottschling 2001). Our results reported here are consistent with that idea: *mre11Δ*-cells have an altered telomeric end-structure, consisting in shorter G-tails (figs. 1, 3). However, Cdc13p-binding to telomeres appears not to be reduced in *mre11Δ*-cells (Tsukamoto et al. 2001), which suggests that *in vivo*, the length requirements for binding of this protein to G-tails may be less than 11 bases, the minimal length required for binding *in vitro* (Hughes et al. 2000). Alternatively, there may be a different mode of Cdc13p-binding on yeast telomeres with very short

overhangs. Taken together, the data suggest that in absence of the MRX-complex, alternative activities establish an altered and shortened G-tail on yeast telomeres. This also suggests that the establishment of such an end-structure and the binding of Cdc13p are essential features for telomere-capping.

These suggested roles for the MRX-complex in processing yeast telomeres may be conserved in mammals. For example, the functional homologue of the yeast Xrs2p in humans, NBS1, associates transiently with a telomeric complex of hRAD50/hMRE11/TRF2 (Zhu et al. 2000). This association occurs in S-phase of the cell cycle and it has been suggested that the association of NBS1 with hRAD50 and hMRE11 at telomeres either is necessary to prepare the telomeres for replication, or for a post-replicative processing event, such as the generation of appropriate G-tails and/or the formation of t-loops (Zhu et al. 2000). Furthermore, there is evidence for an S-phase specific regulatory role for the NBS1 protein in maintaining telomeres in the absence of telomerase in human cell lines (Wu et al. 2000). The MRX-complex is also involved in intra-S checkpoint signalling (D'Amours and Jackson 2001). On telomeres, the DNA-strand replicated by leading strand synthesis is thought to become a transiently blunt-ended molecule. Such ends may be sensed by this checkpoint and processing to expose a single-stranded 3' end could be initiated. Therefore, our results raise the possibility that such leading-strand ends may have particular requirements for recognition and processing, which would involve the MRE11-complex in yeast and humans.

MATERIALS AND METHODS

Yeast cell growth, DNA manipulations and analyses followed standard and published protocols. For details, see the Materials and Methods section in the Supplementary Materials (webpage address).

Primer extension analysis of G-tails. 0.1 ng of the oligo DUP-16 were boiled for 10 min in PE1 buffer (20 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 50 mM NaCl, 1 mM dithiothreitol) and let to anneal at 23°C for more than one hour. Chromosomal terminal restriction fragments (TRFs) were gel purified using a QIAquick gel extraction kit (Qiagen). Standard poly-A tailing reactions were performed with annealed DUP-16 oligo or gel purified TRFs (details to be published elsewhere). Radiolabelled poly-T+(C/A) oligo (5'-CGGAATTCC(T)₁₈M-3'; M = C or A) was annealed to these 3' terminally tailed products in PE1 buffer at 16 °C for 16 hours. Primer-extension reactions were performed on annealed substrates in PE1 buffer, which also contained 0.325 mM dATP, 0.325 mM dCTP and 20 µg BSA, at 12 °C for 20 min. In these conditions, strand displacement occurs on less than 10% of products (not shown and fig.3). Reactions were stopped and loaded on a 15% polyacrylamide gel.

ACKNOWLEDGEMENTS

We thank P. Szankasi for yeast strains and S. Martin for help with the pGTxx plasmids.

We also thank D. Gottschling for a critical reading of the manuscript, anonymous reviewers for comments and suggestions and all members of the Wellinger laboratory for many valuable discussions. This work was supported by a research grant (MOP 12616), and a core group grant (GRP86284) from the Canadian Institutes for Health Research.

M.L. was supported by an MRC studentship, C.L. by a studentship of NSERC and R.J.W. is a Chercheur National of the FRSQ.

REFERENCES

- Blackburn, E.H. 2001. Switching and signaling at the telomere. *Cell* **106**: 661-73.
- Boulton, S.J. and S.P. Jackson. 1998. Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *Embo J* **17**: 1819-28.
- Chakhparonian, M. and R.J. Wellinger. 2003. Telomere maintenance and DNA replication: how closely are these two connected? *Trends Genet* **19**: 439-46.
- D'Amours, D. and S.P. Jackson. 2001. The yeast Xrs2 complex functions in S phase checkpoint regulation. *Genes Dev* **15**: 2238-49.
- . 2002. The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. *Nat Rev Mol Cell Biol* **3**: 317-27.
- Diede, S.J. and D.E. Gottschling. 2001. Exonuclease activity is required for sequence addition and Cdc13p loading at a de novo telomere. *Curr Biol* **11**: 1336-40.
- Dionne, I. and R.J. Wellinger. 1996. Cell cycle-regulated generation of single-stranded G-rich DNA in the absence of telomerase. *Proc Natl Acad Sci U S A* **93**: 13902-7.
- Haber, J.E. 1998. The many interfaces of Mre11. *Cell* **95**: 583-6.
- Hughes, T.R., R.G. Weilbaecher, M. Walterscheid, and V. Lundblad. 2000. Identification of the single-strand telomeric DNA binding domain of the *Saccharomyces cerevisiae* Cdc13 protein. *Proc Natl Acad Sci U S A* **97**: 6457-62.

- Ivanov, E.L., N. Sugawara, C.I. White, F. Fabre, and J.E. Haber. 1994. Mutations in XRS2 and RAD50 delay but do not prevent mating-type switching in *Saccharomyces cerevisiae*. *Mol Cell Biol* **14**: 3414-25.
- Kironmai, K.M. and K. Muniyappa. 1997. Alteration of telomeric sequences and senescence caused by mutations in RAD50 of *Saccharomyces cerevisiae*. *Genes Cells* **2**: 443-55.
- Nugent, C.I., G. Bosco, L.O. Ross, S.K. Evans, A.P. Salinger, J.K. Moore, J.E. Haber, and V. Lundblad. 1998. Telomere maintenance is dependent on activities required for end repair of double-strand breaks. *Curr Biol* **8**: 657-60.
- Tomita, K., A. Matsuura, T. Caspari, A.M. Carr, Y. Akamatsu, H. Iwasaki, K. Mizuno, K. Ohta, M. Uritani, T. Ushimaru, K. Yoshinaga, and M. Ueno. 2003. Competition between the Rad50 complex and the Ku heterodimer reveals a role for Exo1 in processing double-strand breaks but not telomeres. *Mol Cell Biol* **23**: 5186-97.
- Tsukamoto, Y., A.K. Taggart, and V.A. Zakian. 2001. The role of the Mre11-Rad50-Xrs2 complex in telomerase-mediated lengthening of *Saccharomyces cerevisiae* telomeres. *Curr Biol* **11**: 1328-35.
- Wellinger, R.J., K. Ethier, P. Labrecque, and V.A. Zakian. 1996. Evidence for a new step in telomere maintenance. *Cell* **85**: 423-33.

- Wellinger, R.J., A.J. Wolf, and V.A. Zakian. 1993a. Origin activation and formation of single-strand TG1-3 tails occur sequentially in late S phase on a yeast linear plasmid. *Mol Cell Biol* **13**: 4057-65.
- . 1993b. Saccharomyces telomeres acquire single-strand TG1-3 tails late in S phase. *Cell* **72**: 51-60.
- Wu, G., W.H. Lee, and P.L. Chen. 2000. NBS1 and TRF1 colocalize at promyelocytic leukemia bodies during late S/G2 phases in immortalized telomerase-negative cells. *J Biol Chem* **275**: 30618-22.
- Zhu, X.D., B. Kuster, M. Mann, J.H. Petrini, and T. Lange. 2000. Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. *Nat Genet* **25**: 347-52.

FIGURE LEGENDS

Figure 1. G-tail signals are reduced in *mre11Δ* strains.

A) Genomic DNA derived from α -factor arrested cells was digested with *XhoI*. G-tails on YLpFAT10 (indicated by asterisk) were analyzed by native in-gel hybridization for wt and *mre11Δ* strains (MLY50 and MLY51). DNAs were either mock-treated (labelled ExoI-), or treated with *E. coli* Exonuclease I (labelled ExoI+), before in-gel analysis. Top left panel: Ethidium-bromide stained gel; top right panel: the gel after non-denaturing hybridization to the CA-probe; bottom left panel: the DNA was denatured in the gel, transferred to a nylon membrane and the membrane hybridized to a Y'-probe. Single-stranded phagemid DNA containing yeast telomeric repeats of the G-rich strand (ssGT) and of the C-rich strand (ssCA) served as a positive and negative controls respectively. The ssGT control was mixed with *PvuI*-digested pMW55, the latter serving as double-stranded control (ds, a fragment of 1.9 kb). Linearized pVZY'K plasmid (3.9 kb) served as a positive control for the hybridization with the Y' probe (labelled Y'). M: end-labelled 1 kb ladder DNA serving as a size standard.

B) Relative amounts for G-tails (relative signals for ssDNA, in arbitrary units) were obtained by calculating the ratio of the respective signals on the native gels over the total amount of Y' TRF signals or YLpFAT10 signals. Three to six individual non-synchronous cultures were tested for each strain and the standard deviation is indicated.

Figure 2. Most G-tails on telomeres in G1-arrested cells are shorter than 16 bases.

A) The individual pGTxx plasmids, each containing telomeric repeat tracts ranging from 10 to 22 bp, were double-digested and heat denatured prior to loading on an agarose gel. A common DNA fragment serves as loading control after hybridization to an M13-probe (open arrowheads), while a fragment migrating at about 1.3 kb of each plasmid contains the TG₁₋₃ repeats of varying lengths serving to compare the hybridization of the CA-probe with different telomeric repeat lengths (closed arrowheads). Wild-type cells (MLY30 strain) were arrested in G1 with α -factor and *XhoI*-digested DNAs derived from those cells were loaded on the gel (lane marked WT). After hybridization to the M13-probe (top left), the same gel was then hybridized at 4°C to the CA-probe and washed at different temperatures as indicated at the bottom of gels shown. The signals obtained for G-tails on YLpFAT10 (indicated by asterisk) were compared to those detected on the 1.3 kb DNA fragments derived from the pGTxx plasmids after different washing temperatures. Controls and molecular size standards are as described in fig. 1.

B) Remaining signals on pGTxx fragments as well as signals for G-tails on YLpFAT10 were determined by using native gels washed at different temperatures. Signals are plotted as % with respect to the same gel washed at 4°C (designated at 100%).

Figure 3. Constitutive G-tails are shorter in *mre11Δ* cells.

Top: DNA structure of the annealed model substrate DUP-16 and expected structure for TRFs. In DUP-16, the two nucleotides not corresponding to telomeric sequence and where strand displacement synthesis stops are indicated in red. Bottom: analysis of extension products by polyacrylamide gel electrophoresis. DUP-16 and/or TRFs were poly-A tailed, annealed to end-labelled poly-T+(C/A) oligo and subjected to an extension

reaction using T4 DNA polymerase. Left gel: control reactions using DUP-16 alone; w/o dCTP: dCTP was omitted from the extension reaction. Poly-T+(C/A) indicates the position of the end labelled oligo alone and corresponds to position + 1 in terms of overhang size. Polymerized product: product of the extension reaction on DUP-16 covering the overhang only (+ 24); displacement product: product of the extension reaction reaching the 2 inversed G:C base pairs (+ 31). Right gel: purified TRFs from the indicated strains were subjected to the procedure. In the last two lanes, either DUP-16 or purified TRFs from a wt strain were treated with ExoI before the tailing reaction and then processed as the other samples. Size indications on right are with respect to overhang size.

Figure 4. Most yeast telomeres end with a G-tail.

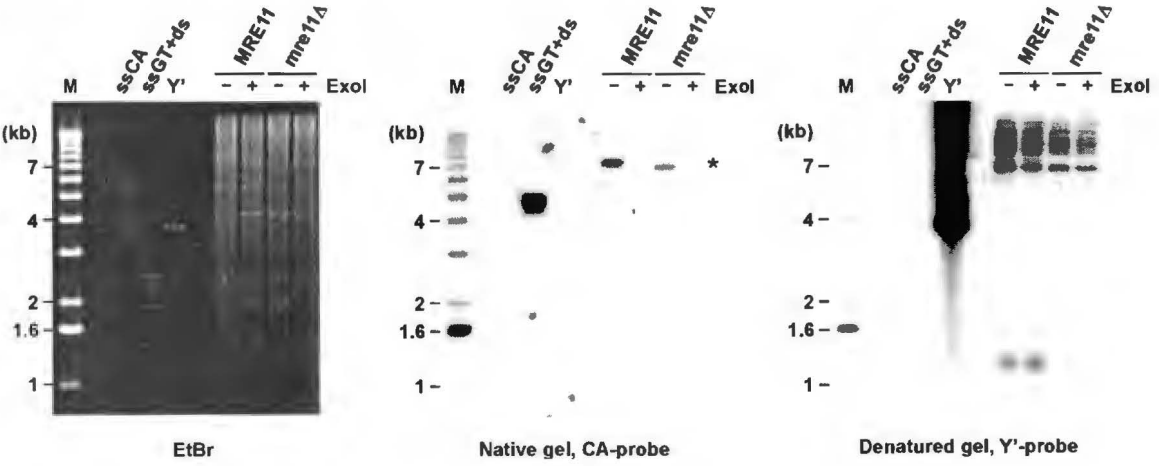
A) Upper panel: genomic DNAs isolated from the indicated strains were first treated with Exonuclease I (labelled blunted) or mock treated (labelled native). DNAs were then mock-treated (labelled ExoIII -), or treated with Exonuclease III (labelled ExoIII +), before native gel analysis. Closed arrowhead: internal positive control for the ExoIII treatment (*Acc65I*-digested pRS304); open arrowhead: internal negative control (*KpnI*-digested pRS314), revealed by the T7-probe. *KpnI*-digested TRFs are seen with the GT-probe only in lanes where genomic DNA was blunted with ExoI prior to ExoIII reaction. Lower panel: same gel as above after denaturation and rehybridization to a 2 µm probe to show equal loading of DNA (indicated by asterisk). Controls and molecular size standards are as in fig. 1.

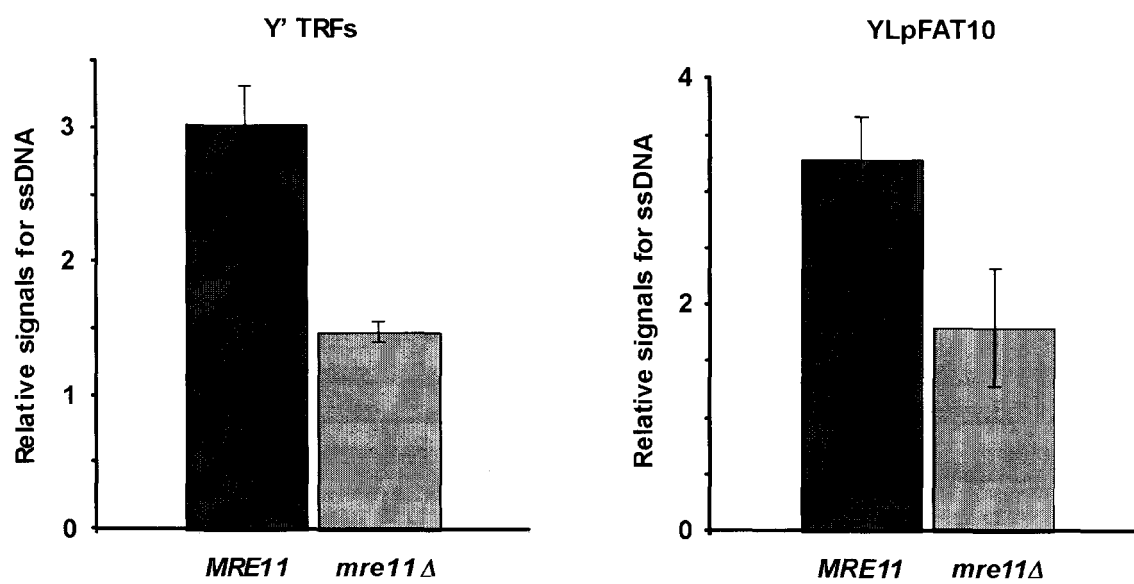
B) Percentage of telomeres ending with a G-tail was estimated by quantifying the amount of G-tail signals found on YLpFAT10 on native gels washed at 4°C (see fig. 2A). The values obtained were compared with signals detected for pGT19 and pGT22 fragments and were adjusted for the molecular ratio of hybridization targets on pGT19/pGT22 vs the YLpFAT10 plasmid (see Suppl. Materials). DNAs derived from three independent cultures of wt cells were used for quantification and the standard deviation is indicated.

Figure 5. S-phase dependent G-tail formation in *mre11Δ*-cells.

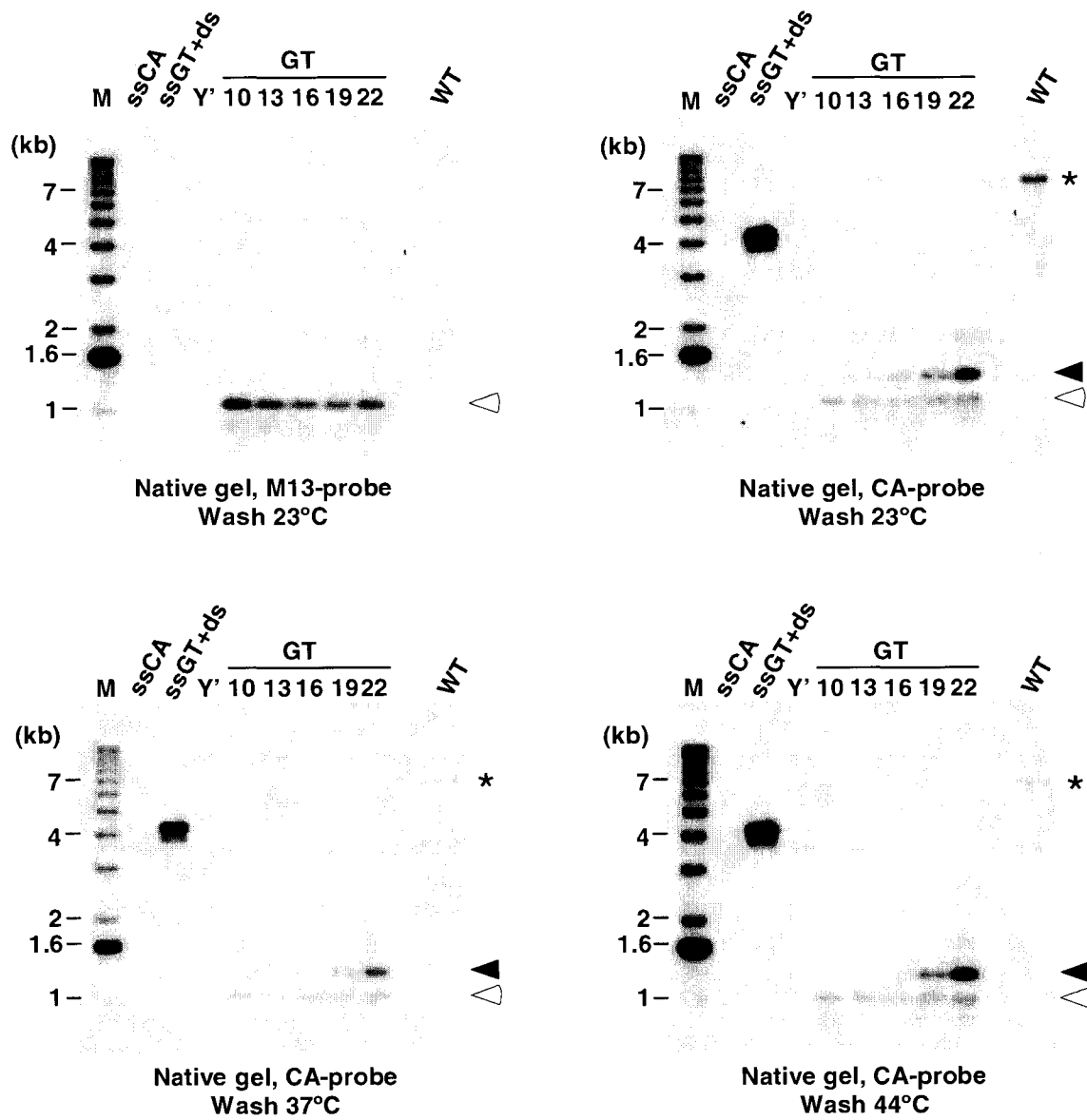
A cell synchronization experiment using *MRE11* and *mre11Δ* cells (AR120 and MLY40 strains) was performed as described in supplementary fig. 1. DNA isolated from each cell aliquot was analyzed for telomeric DNA end-structures by non-denaturing in-gel hybridization (Supplementary fig.2 and data not shown). The relative amounts for telomeric ssDNA on YLpFAT10 (in arbitrary units) were obtained by calculating the ratio of the G-tail signals retrieved on the native gel over the total amount of YLpFAT10 detected.

Larrivée, M., LeBel, C. and Wellinger, R.J.
Figure 1A

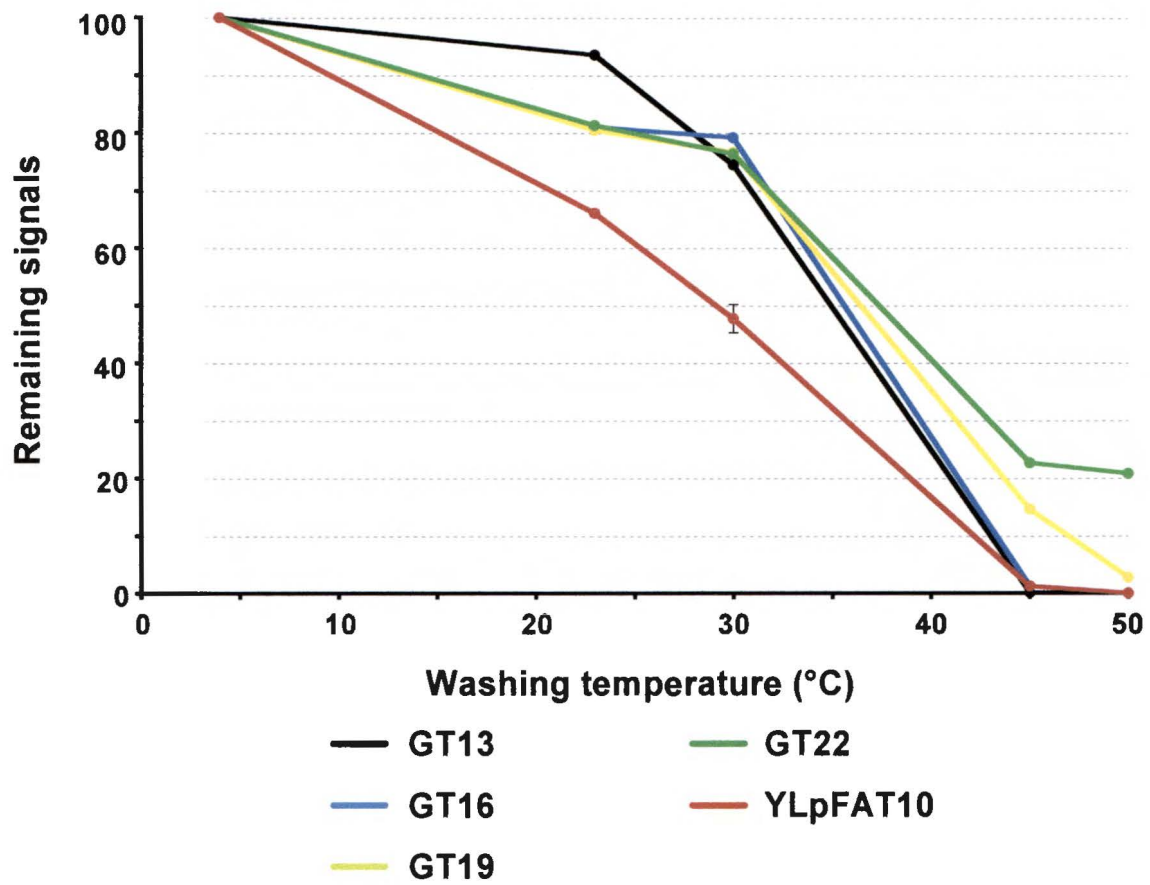




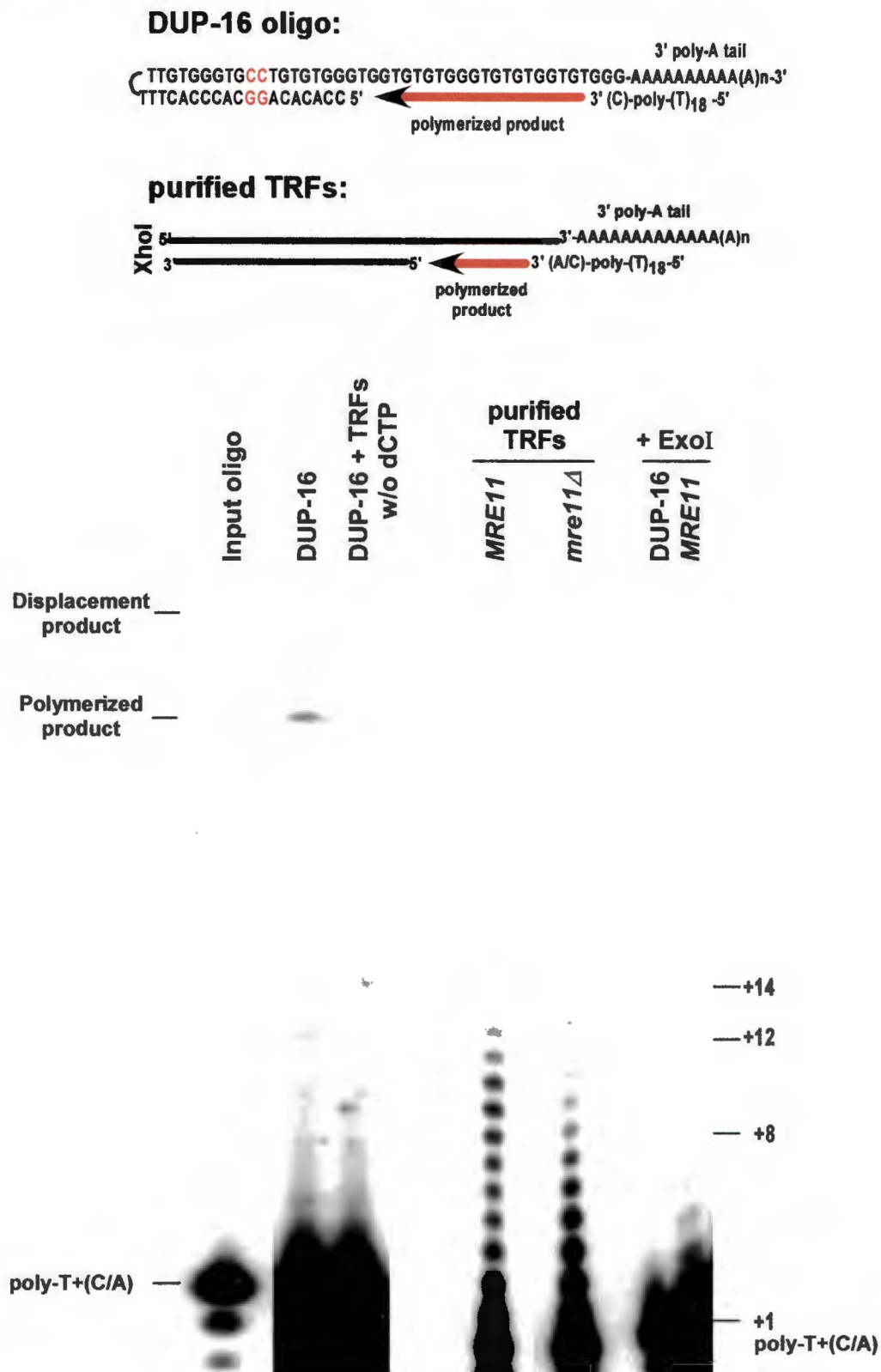
Larrivée, M., LeBel, C. and Wellinger, R.J.
Figure 2A



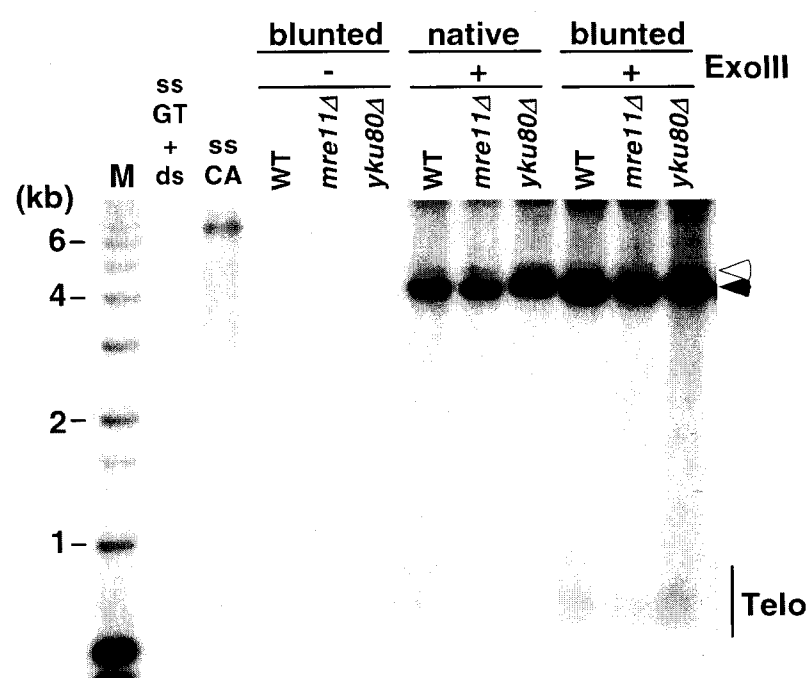
Larrivée, M., LeBel, C. and Wellinger, R.J.
Figure 2B



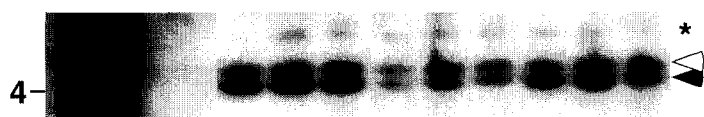
Larrivée, M., LeBel, C. and Wellinger, R.J.
Figure 3



Larrivée, M., LeBel, C. and Wellinger, R.J.
 Figure 4A

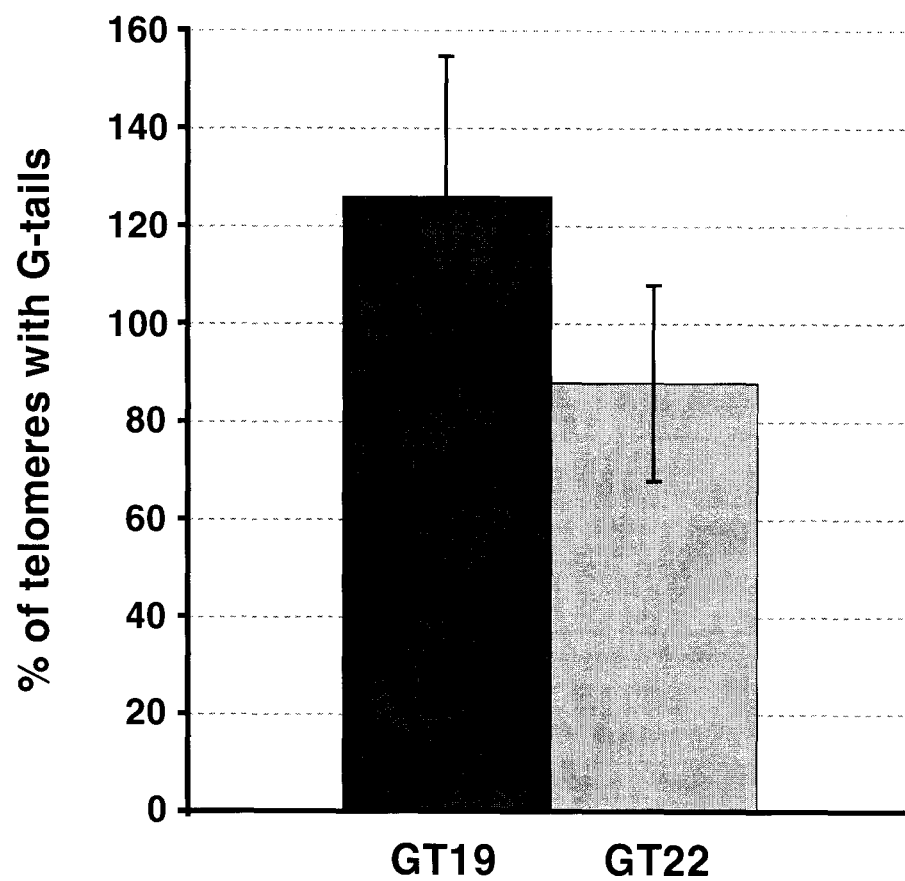


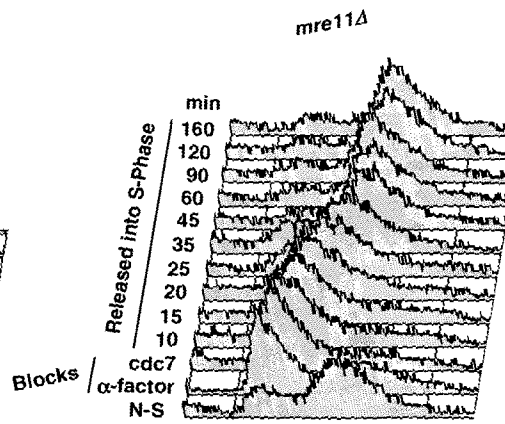
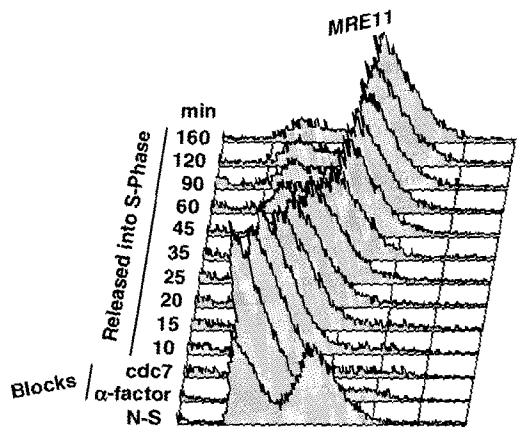
Native gel, GT-probe and T7-probe



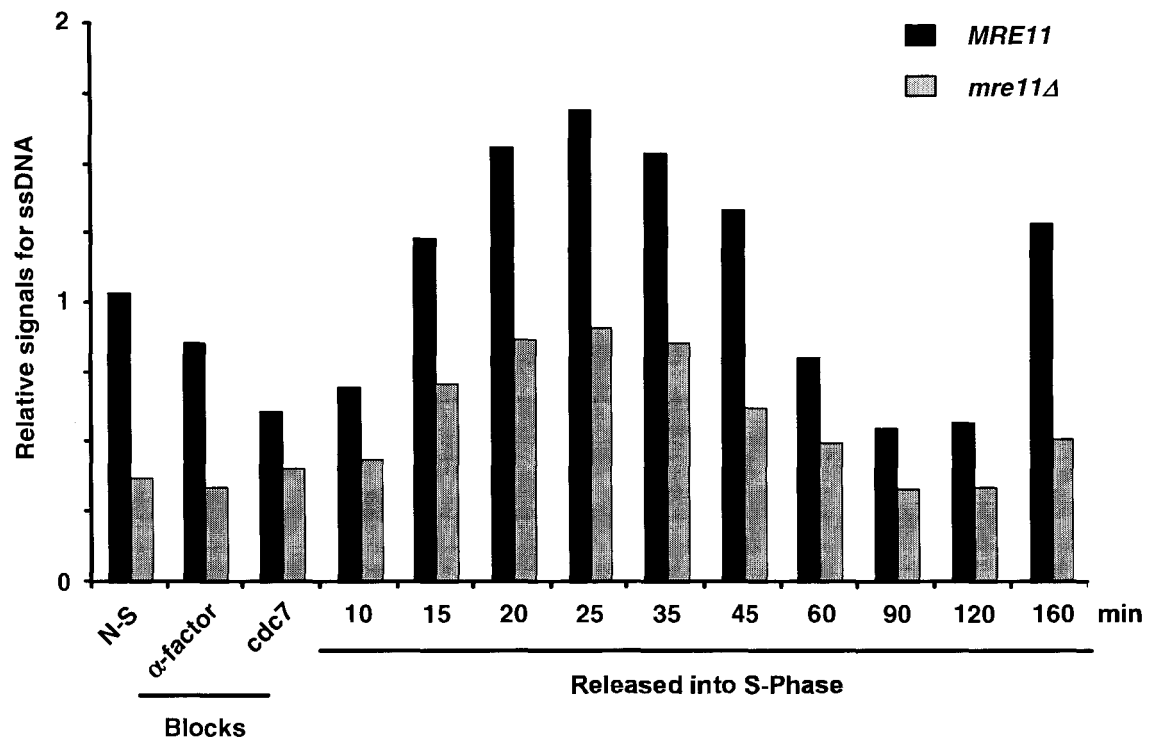
Denatured gel, 2μm-probe

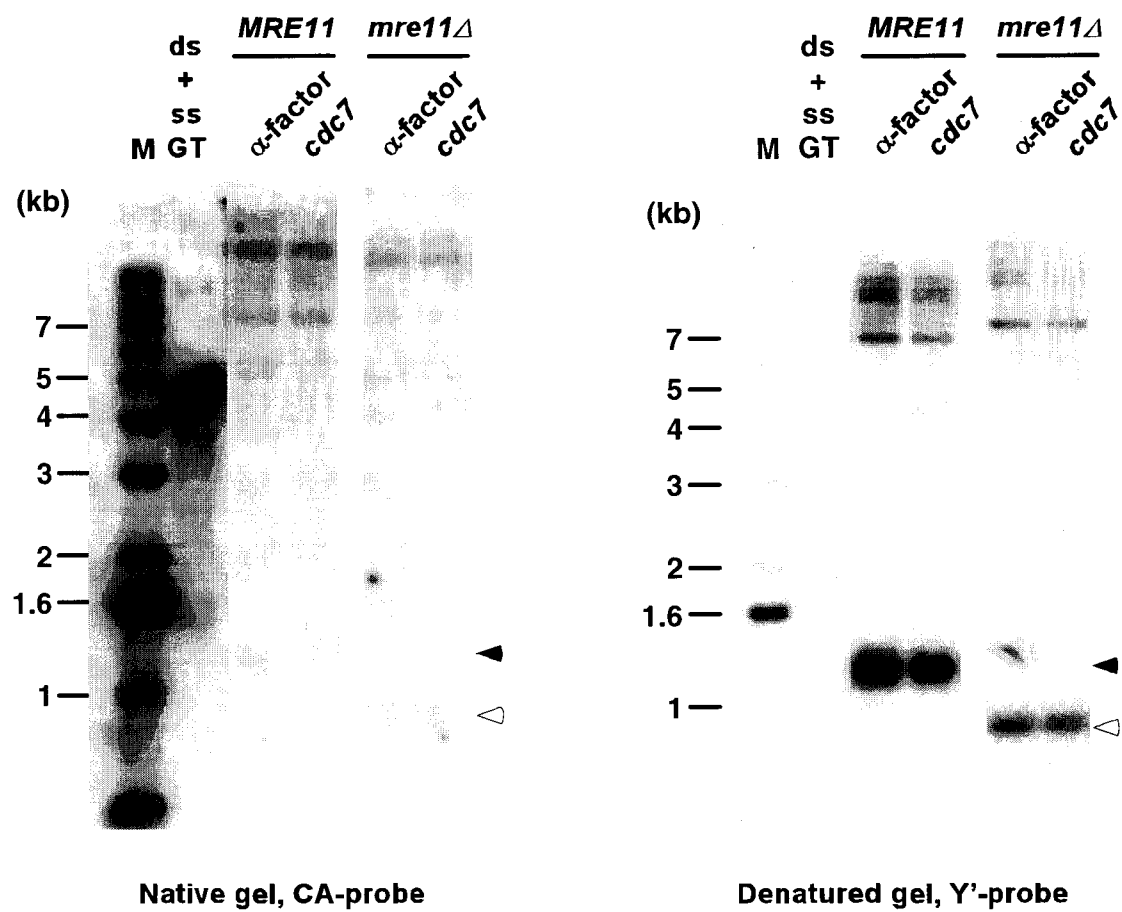
Larrivée, M., LeBel, C. and Wellinger, R.J.
Figure 4B





Larrivée, M., LeBel, C. and Wellinger, R.J.
Figure 5B





CHAPITRE 4

Analyse de la longueur des extensions 3' télomériques de différentes lignées cellulaires humaines.

Préambule

Selon le modèle de protection des télomères proposé par E.H. Blackburn (Blackburn 2000), l'intégrité de l'extension 3' simple-brin joue un rôle majeur dans la formation du capuchon télomérique et dans la protection des extrémités des chromosomes. En effet, de récentes observations montrent que ce sont des changements dans certains aspects de l'état des télomères plutôt que la taille globale des télomères *per se* qui activeraient la sénescence réplivative (Karlseder et al. 1999; van Steensel et al. 1998; Takai et al. 2003). Il a été montré que lors de la production d'une forme dominante négative de la protéine TRF2 dans les cellules, il y a perte partielle de l'extension 3' simple-brin et déprotection des télomères, ce qui peut mener à des fusions entre télomères (van Steensel et al. 1998; Karlseder et al. 2002).

Afin d'étudier l'importance de l'extension télomérique chez les cellules humaines, et en particulier chez des cellules maintenant la longueur des télomères par un mécanisme autre que l'activation de l'enzyme télomérase (cellules ALT) (Bryan and Reddel 1997; revue dans Henson et al. 2002), nous avons apporté des modifications à une méthode déjà publiée nommée T-OLA (Telomeric Oligonucleotide Ligation Assay) (Cimino-Reale et al. 2003). Dans cette section, je décris les différentes étapes de mise au point, faites en premier lieu sur des oligonucléotides, afin de valider notre nouvelle approche expérimentale. Je décris également les expériences préliminaires de T-OLA

réalisées sur l'ADN de cellules normales, transformées, cancéreuses ou ALT, visant à caractériser l'extension 3' télomérique. Étant donné le peu de méthodes précises permettant l'analyse en détails de l'extrémité 3', l'approche mise au point visait à augmenter la précision dans la mesure de l'extension simple-brin grâce à des modifications apportées à la méthode de base. Sachant maintenant le rôle crucial de l'extension 3' simple-brin dans la protection des télomères, celle-ci doit être caractérisée avec soin afin de mieux comprendre sa dynamique et son implication dans différents processus cellulaires. J'ai effectué toutes les expériences de cette section. Les mises au point concernant cette méthode et les essais de mesure des extensions télomériques avec cette méthode sont présentement poursuivis par un étudiant du laboratoire, et pourraient éventuellement mener à un manuscrit.

MATÉRIEL ET MÉTHODES

Lignées cellulaires, extraction de l'ADN des cellules humaines, traitement par des exonucléases et digestions enzymatiques

Les différentes lignées cellulaires utilisées dans ces essais ont été mises en culture et maintenues en croissance dans plusieurs laboratoires. Des aliquots ou des culots de cellules nous ont été fournis pour nos expériences (dons des laboratoires du Projet MoNA/Sherbrooke, Télogène Inc. et du Dre. Chantal Autexier, Université McGill). L'ADN de ces différentes lignées cellulaires a été obtenu grâce à un protocole de lyse cellulaire standard. Brièvement, les cellules confluentes sont récupérées d'un pétri de 60 mm ($\sim 5 \times 10^5$ cellules) et rincées avec 5 ml de PBS A froid (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4 , 1.8 mM KH_2PO_4 , pH 6.8). Par la suite, 1ml de PBS A froid est déposé sur les cellules. Les cellules sont décollées du pétri à l'aide d'un policeman stérile et la suspension cellulaire est transférée dans un tube de 1.5 ml. Le culot cellulaire est récupéré suite à une courte centrifugation. Ensuite, 500 μl de tampon DNA-A (10 mM Tris, pH 7.4, 10 mM EDTA, 10 mM NaCl) et 2.5 μl de RNase A 20 mg/ml sont ajoutés au culot et celui-ci est incubé 1 heure à 37°C. La protéinase K est ajoutée à une concentration finale de 100 $\mu\text{l}/\text{ml}$ et du SDS est rajouté à une concentration finale de 1%. Le tube est inversé doucement pour mélanger et incubé 3 heures à 37°C avec agitation douce. Par la suite, l'ADN subit 3 extractions avec du phénol-chloroforme puis est précipité par l'ajout de 0.1 M NaCl et 2 volumes d'éthanol à température ambiante. Le tube est inversé plusieurs fois jusqu'à l'obtention d'un précipité. Celui-ci est rincé avec de l'éthanol 70% (v/v) et le culot est séché à l'air libre, puis resuspendu dans 500 μl de

TE pH 8.0, à 37°C pour 12 heures. La provenance et le type de lignées cellulaires utilisées pour nos expériences sont résumés dans le Tableau 7.

L'ADN cellulaire purifié (5-10 µg au total) a été traité par la Mung Bean Nuclease (Amersham Biosciences; 147.2 U/µl) ou par l'exonucléase I de *E. coli* (USB; 10 U/µl), selon les recommandations du fabricant.

Les digestions enzymatiques (EcoRV, HinfI/RsaI; New England Biolabs) effectuées sur l'ADN purifié de cellules humaines (5-10 µg au total) ont été faites selon les recommandations du fabricant.

Lignées cellulaires	Caractéristiques des lignées cellulaires	Provenance
HeLa S3	Cellules épithéliales dérivées de la lignée parentale de HeLa, adénocarcinome, expriment HPV 18	Télogène Inc.
ECR 293	Cellules épithéliales embryonnaires de rein, transformées par adénovirus 5	Projet MoNA
HCT 116	Cellules épithéliales provenant d'un carcinome colorectal, produisent des tumeurs chez des souris immunodéficientes	Projet MoNA
HA 5	Cellules embryonnaires de rein, transformées avec SV 40	Dre Autexier (McGill)
BJ	Fibroblastes provenant de prépuce, possèdent une longue vie en comparaison avec d'autres lignées de fibroblastes (85-90 PD), sont négatives pour l'expression de la télomérase	Projet MoNA
WI 38	Fibroblastes embryonnaires provenant de poumons normaux, peuvent faire 50 ±10 PD	Dre Autexier (McGill)
GM 847	Fibroblastes de peau, maintiennent les télomères par processus ALT, négatives pour l'expression de la télomérase, transformées par SV40	Dre Autexier (McGill)(Hwang et Kucherlapati 1983; Bryan et Reddel 1997)

Tableau 7. Caractéristiques des lignées cellulaires étudiées.

Oligonucléotides et réaction d'échange de phosphate

L'oligo TS10 (aussi appelé substrat TS10) de séquence 5'-TTAGGGTTAGGG-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG-3' utilisé dans les essais de mise au point a été obtenu de Gibco BRL.

L'oligo phosphorylé P-C₃TA₂ purifié par HPLC (aussi appelé ligand C) de séquence 5'-P-CCCTAA-3', a été obtenu de Integrated DNA Technologies Inc. Typiquement, 40-50 pmoles de l'oligo P-C₃TA₂ étaient utilisées par réaction d'échange de phosphate radiomarké avec 1,2 X 10⁻⁴ Ci de γ -ATP³², selon le protocole de réaction d'échange de Sambrook et al. (1989).

Le plasmide pRST5 contient 576 pb de séquences télomériques humaines (~ 96 répétitions télomériques TTAGGG) (Griffith et al. 1999). Ces séquences télomériques ont été clonées au site BbsI du plasmide pRST15, qui est basé sur le plasmide pBluescript SK⁺. Les séquences télomériques purifiées de ce plasmide sont utilisées comme contrôle pour la réaction de polymérisation. Ce plasmide provient du laboratoire du Dr. J. Griffith.

Le plasmide pTH5 (de Lange et al. 1990) contient 200 pb de séquences télomériques humaines, qui ont été clonées au site EcoRI du plasmide pSP73. Les séquences télomériques purifiées à partir de ce plasmide sont utilisées comme contrôle pour la réaction de polymérisation. Ce plasmide provient du laboratoire du Dre. T. de Lange.

Les plasmides contenant les séquences télomériques sont d'abord digérés avec les enzymes de restriction appropriées, selon les recommandations du manufacturier. Les fragments générés par les digestions enzymatiques sont séparés sur un gel 0,6% d'agarose dans un tampon TAE. Les fragments dont la taille correspond aux digestions

des séquences télomériques sont découpés du gel. Le fragment d'ADN contenu dans l'agarose est purifié sur des colonnes QIAquick permettant la purification de l'ADN à partir d'un gel d'agarose (QIAquick Gel Extraction), selon le protocole du fabricant. L'ADN total purifié est par la suite dosé sur gel afin d'estimer sa concentration totale.

Conditions d'hybridation, de ligation et d'électrophorèse utilisées lors des essais de mesure du simple-brin (adaptation du protocole de base publié par (Cimino-Reale et al. 2001)) par la méthode de T-OLA modifiée

Dans les essais de mise au point avec le substrat TS10 ou les substrats de longue taille (pRST5 et pTH5), 1 ou 10 pg d'oligo étaient chauffées à 100°C pour 10 minutes en présence de 2 pmoles de γ -³²P-C₃TA₂, dans un tampon de ligation (50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 10 mM DTT, 1mM ATP, 25 µg/ml BSA). Les tubes étaient incubés à 12°C pour environ 16 heures pour permettre l'hybridation des oligos. 400 unités de T4 ADN ligase (400 000 U/ml; NEB) étaient ajoutées à chaque tube de réaction, puis les tubes étaient incubés à 12°C pour 5 heures. Les produits de réaction subissaient une extraction avec du phénol-chloroforme et une précipitation à l'éthanol. Les culots récupérés étaient resuspendus dans 5 µl de tampon formamide et mis sur gel dénaturant 15% de polyacrylamide (pré-migré pour 1 heure à 40 W en présence d'une plaque de métal) puis migré pour 2h30 à 40 W avec une plaque de métal.

Pour les réactions contenant l'ADN génomique purifié de différentes lignées cellulaires, environ 10 µg d'ADN étaient utilisés. Les réactions étaient effectuées comme décrites auparavant. Les échantillons étaient chargés sur un gel 5% polyacrylamide dénaturant, et migrés dans les mêmes conditions.

RÉSULTATS

Des modifications apportées à la méthode de T-OLA permettent la mesure de la longueur d'oligos télomériques contrôlés

Afin de mieux comprendre l'importance et la dynamique des extensions 3' simple-brin chez différentes lignées de cellules humaines, nous avons apporté des modifications à une méthode déjà existante (T-OLA) (Cimino-Reale et al. 2003), dans le but d'augmenter la précision dans la mesure de l'extension 3' simple-brin. La méthode est basée sur l'hybridation d'oligos complémentaires (CCCTAA)_n radiomarqués aux extensions télomériques humaines, et la ligation subséquente des oligos hybridés. Avec cette méthode, il est possible d'avoir un aperçu de la longueur totale de l'extension simple-brin, par la taille du produit de ligation des oligos complémentaires à leur substrat (figure 27).

Les différentes étapes de la méthode de T-OLA sont décrites dans la figure 27. La méthode de T-OLA telle que mise au point et utilisée dans certaines publications permet la mesure des extensions télomériques simple-brin auxquelles ont été hybridés les oligos (CCCTAA)_n (Cimino-Reale et al. 2001; Stewart et al. 2003). Lors de la mise au point de la méthode, la longueur des oligonucléotides utilisés pour l'hybridation variait entre 12 et 24 nucléotides. Les températures d'hybridation varient aussi selon la longueur de l'oligo utilisé (30°C pour le 12-mer, 50°C pour le 18-mer et 65°C pour le 24-mer) (Cimino-Reale et al. 2001). La réaction de ligation des différents oligos hybridés au substrat se fait à la même température que la réaction d'hybridation, avec une enzyme thermostable (Cimino-Reale et al. 2001).

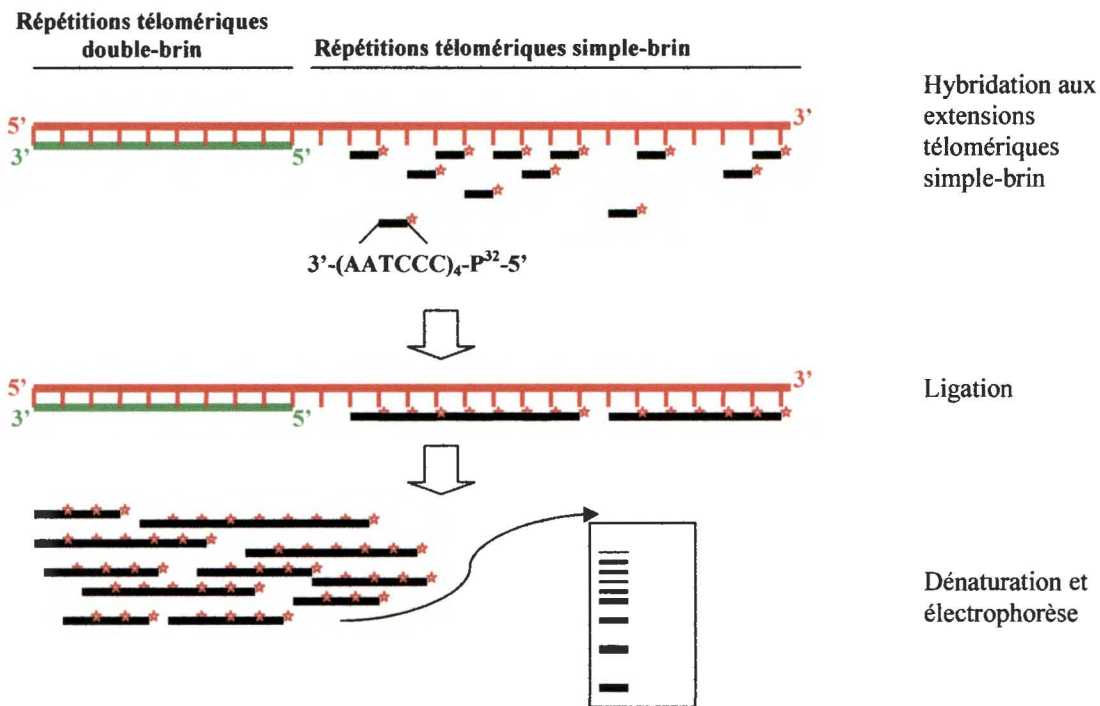


Figure 27. Schéma des étapes de la réaction classique de T-OLA. L'ADN génomique total purifié des différentes lignées cellulaires est incubé avec un oligonucléotide contenant un phosphate radiomarké en 5', dont la séquence est (CCCTAA)₄ (fragment noir avec étoile rouge). Cet oligonucléotide s'hybride aux extensions télomériques simple-brin qui sont accessibles. La ligase est ajoutée à cette réaction, permettant la ligation des oligonucléotides hybridés à des positions adjacentes sur l'extension simple-brin. Cette approche permet de générer différents fragments d'ADN radiomarqués, dont les tailles augmentent de 24 nucléotides à la fois. Ces fragments sont dénaturés et séparés sur un gel dénaturant de polyacrylamide.

La reproductibilité et l'exactitude des résultats obtenus avec cette méthode dépendent de l'accomplissement total de plusieurs réactions. Plusieurs étapes dans la méthode peuvent être incomplètes et nuire à l'obtention des résultats valides. Premièrement, des oligos (CCCTAA)₄ qui n'ont pas été préalablement phosphorylés peuvent se retrouver dans la réaction. Ceux-ci peuvent facilement s'hybrider au substrat, mais ne peuvent pas être ligués aux autres oligos dû au manque de phosphate en 5'. Cette condition peut mener à une accumulation finale de courts produits de ligation. Deuxièmement, la taille des oligos utilisés peut aussi empêcher l'hybridation d'autres oligos sur le substrat. En effet, l'hybridation des oligos sur l'extension n'est pas un

processus coopératif, i.e. l'hybridation d'un oligo à un site optimal n'entraîne pas l'hybridation d'un autre oligo au site optimal situé juste à côté de lui. L'hybridation aléatoire des oligos sur le substrat peut donc mener à l'apparition de « trous d'hybridation », où aucun oligo n'est apparié à l'extension. Les trous présents entre deux oligos consécutifs hybridés sur le substrat peuvent être de 6, 12, 18, 24, etc. nucléotides. De ces différentes tailles possibles, seuls les trous de 24 nucléotides et plus pourront servir de substrat à un oligonucléotide, et conséquemment être hybridés par un oligo. Les trous de taille inférieure à 24 nucléotides ne seront pas couverts par un oligo sur l'extension télomérique. La présence de ces trous dispersés sur l'extension peut diminuer la ligation continue des oligos, et mener à une accumulation de courts produits de ligation. Une accumulation de courts produits de ligation peut aussi être observée si les réactions d'hybridation et de ligation ne sont pas complètes. Les réactions de T-OLA ont montré que les tailles des extensions télomériques de différentes lignées cellulaires varient de 24 à plus de 400 nucléotides (Cimino-Reale et al. 2001). Dans les fibroblastes et les lymphocytes, il a été montré qu'environ 56% des extensions simple-brin sont de 90 nucléotides et moins (Cimino-Reale et al. 2001). Cette courte extension qui serait présente sur plus de la moitié des télomères est de taille limite pour permettre la formation de t-loops. Il serait donc possible que l'estimation de la longueur des extensions soit le reflet d'un biais de la méthode, introduit par la méthode même de mesure des extensions.

Nous avons apporté quelques modifications de base à la technique de T-OLA, afin vérifier l'exactitude des résultats obtenus et pousser la précision de la mesure des extensions simple-brin. Premièrement, nous avons utilisé un oligo complémentaire aux

extensions télomériques humaines de 6 nucléotides (CCCTAA) plutôt qu'un oligo de 24 nucléotides. L'utilisation d'un oligo de 6 nucléotides peut diminuer les « trous d'hybridation » étant donné que chaque « trou » possiblement présent sur l'extension télomérique peut être hybridé par un oligo. Deuxièmement, cet oligo CCCTAA (appelé ligand C) est aussi pré-phosphorylé en 5' et a été purifié par HPLC. Cette modification vise à ce que la vaste majorité des oligos présents dans la réaction possède un phosphate (radiomarké ou non) à l'extrémité 5'. Lors de la réaction de ligation, tous les oligos hybridés sur l'extension peuvent être ligués. Troisièmement, l'hybridation et la ligation de cet oligo se font à une température de 12°C plutôt que 55°C. Une température d'hybridation et de ligation plus basse peut diminuer les possibilités de désappariement entre les oligos hybridés et les extensions télomériques.

La substitution des oligos utilisés précédemment dans le T-OLA par le ligand C de 6 nucléotides a été vérifiée. Le ligand C pré-phosphorylé peut efficacement acquérir un phosphate radiomarké en 5', suite à une réaction de transfert de phosphate radiomarké avec la T4 PNK (donnée non-montrée). Pour vérifier si l'approche de T-OLA modifiée est possible, différentes conditions de réaction ont été mises au point sur le substrat TS10. Le substrat TS10 est un oligo de 60 nucléotides qui comporte 10 cibles parfaitement complémentaires au ligand C.

Dans un premier temps, différentes conditions pour l'hybridation et la ligation du ligand C sur le substrat TS10 ont été testées (Figure 28 A, pistes 1 à 3). Ces mêmes conditions ont été appliquées sur différents substrats contrôles (oligo (C₃TA₂)₆, brin C-riche de répétitions télomériques humaines ou aucun oligo substrat) afin de vérifier la spécificité de la réaction observée entre le substrat TS10 et le ligand C utilisé (figure

28A, pistes 4 à 9). Afin de vérifier que la réaction de ligation observée est dépendante de la présence de la T4 ADN ligase ajoutée à la réaction, les différentes réactions avec le substrat TS10 ont été incubées en absence de l'enzyme (figure 28A, pistes 10 à 12). Les réactions de T-OLA ne sont possibles que lorsque le substrat utilisé contient les répétitions télomériques humaines du brin G-riche (TTAGGG). Ces réactions sont possibles sur le substrat TS10 à différentes températures (4°C, 12°C et 23°C). La ligation entre eux des différents oligos ligand C hybridés sur le substrat TS10 est absolument dépendante de la présence de l'enzyme T4 ADN ligase. Aucun produit de réaction n'est obtenu lorsque l'on garde les conditions de réaction utilisées précédemment et que l'on omet la ligase dans la réaction (figure 28A, pistes 1 à 3 à comparer avec pistes 10 à 12). Les réactions de T-OLA ont aussi été essayées en incubant le substrat TS10 et le ligand C à différentes températures en présence de la ligase dès le début de la réaction d'hybridation; l'efficacité globale de la réaction était diminuée dans ces conditions (données non-montrées) et ces conditions n'ont plus été considérées.

Par la suite, des diminutions dans la concentration du substrat TS10 ont été testées. Ces diminutions visaient à détecter un ratio minimal entre le ligand C et le nombre de cibles possibles pour l'hybridation, afin d'obtenir une réaction de ligation efficace et un produit final d'une longueur de 60 nucléotides. Les mêmes conditions ont aussi été essayées à différentes températures (12°C, 23°C et 30°C) afin de vérifier leur efficacité (figure 28B).

Les réactions de T-OLA permettent d'obtenir des produits correspondant à la ligation d'un nombre accru de ligand C, soient des produits de 18, 24, 30, 36, 42, 48, 54

et 60 nucléotides. Cependant, il semble que la concentration initiale du substrat ait un effet sur l'obtention des produits de ligation et leur visualisation sur autoradiogramme.

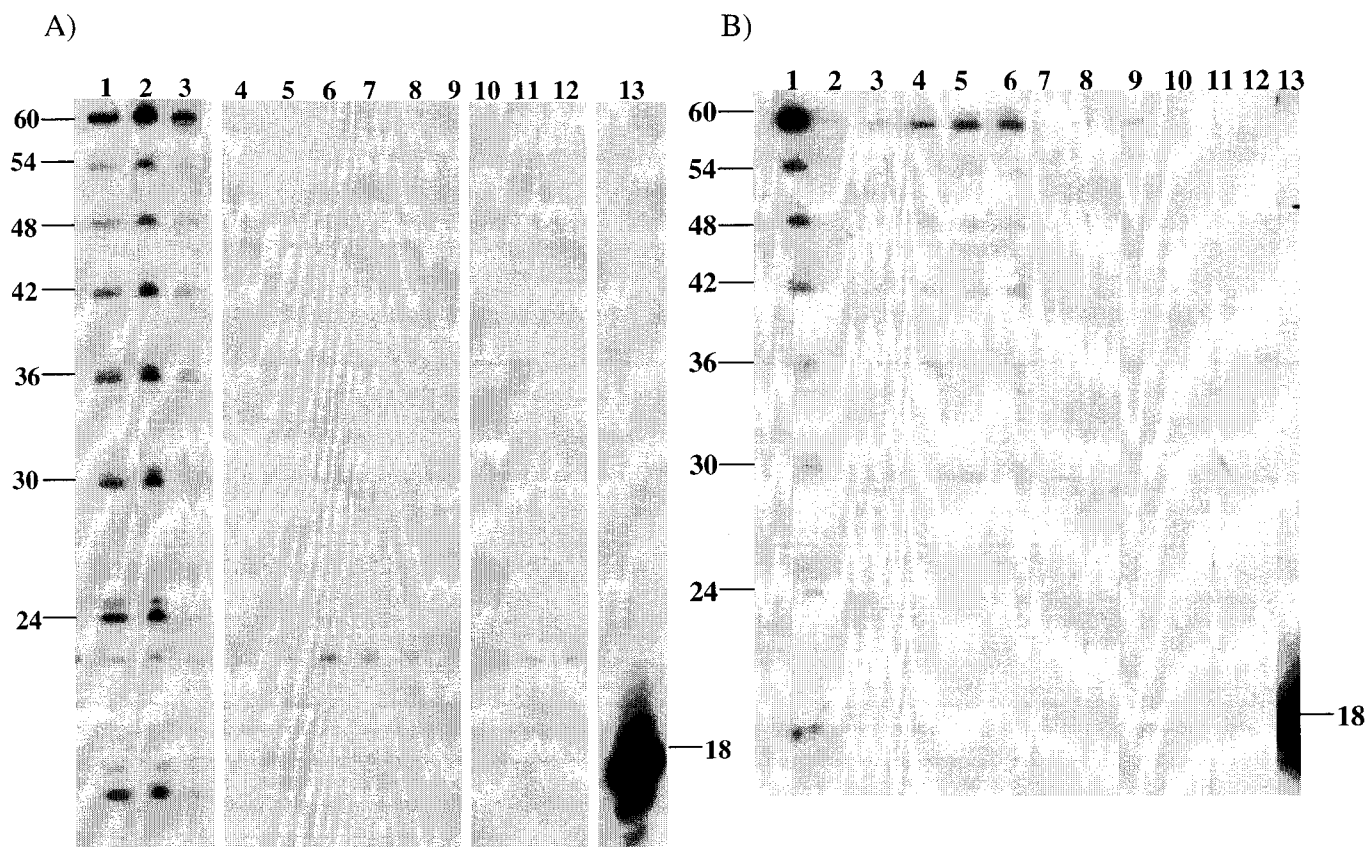


Figure 28. Réactions de T-OLA modifiées; différentes conditions de réaction et différents substrats. A) Effet des températures d'hybridation et de ligation sur la réaction de T-OLA. Pistes 1, 2, 3 : 10 pg de substrat TS10 hybridés et ligués à 4°C, 12°C et 23°C, respectivement. Pistes 4, 5, 6 : 10 pg du substrat (C₃TA₂)₆ hybridés et ligués à 4°C, 12°C et 23°C, respectivement. Pistes 7, 8, 9 : aucun oligo substrat et réaction à 4°C, 12°C et 23°C, respectivement. Pistes 10, 11, 12 : 10 pg de substrat TS10 hybridés et ligués à 4°C, 12°C et 23°C, respectivement, sans l'enzyme T4 ADN ligase. Piste 13 : contrôle de taille, oligo (C₃TA₂)₃ radiomarqué en 5'. La taille des différents produits de ligation est indiquée à gauche de la figure. La bande observée autour de 20-21 nucléotides est considérée comme du bruit de fond causé par l'oligo radiomarqué, et est présente dans toutes les pistes. B) Effet de la température et de la concentration du substrat TS10 sur l'efficacité de la réaction de T-OLA. Pistes 1, 2, 3 : 10 pg de substrat TS10 hybridés et ligués à 12°C, 23°C et 30°C, respectivement. Pistes 4, 5, 6 : 1 pg de substrat TS10 hybridés et ligués à 12°C, 23°C et 30°C, respectivement. Pistes 7, 8, 9 : 0,1 pg de substrat TS10 hybridés et ligués à 12°C, 23°C et 30°C, respectivement. Pistes 10, 11, 12 : 10 pg de substrat (C₃TA₂)₆ hybridés et ligués à 12°C, 23°C et 30°C, respectivement. Piste 13 : contrôle de taille, oligo (C₃TA₂)₃ radiomarqué en 5'. La taille des différents produits de ligation est indiquée à gauche de la figure.

En gardant les conditions optimales mises au point précédemment avec le substrat TS10, les réactions de ligation ont été optimisées afin de récupérer un maximum de produit de 60 nucléotides par rapport aux autres produits de ligation récupérés. Dans cette optique, le rajout d'un excès de ligand C non-radiomarqué après l'étape de ligation a été essayé. Le but visé est de couvrir tous les sites possibles où le ligand C peut s'hybrider,

afin d'optimiser la ligation des oligos ligand C hybridés à tous les sites sur le substrat TS10. L'ajout d'un excès de ligand C phosphorylé en 5' mais non-radiomarké diminue l'intensité du signal total récupéré pour les différents produits de ligation (données non-montrées). De plus, avec ces essais, les produits de ligation récupérés ne sont pas uniquement des produits de 60 nucléotides. La grande diminution dans l'intensité des signaux récupérés suite à ces essais suggère que le rajout d'un excès de ligand C crée un nouvel équilibre dans la réaction d'hybridation. Ce nouvel équilibre favorise le déplacement des molécules de ligand C radiomarquées qui étaient hybridées au substrat TS10 avant le rajout.

Les différentes méthodes de mesure du simple-brin télomérique ont permis d'estimer la taille des extensions télomériques simple-brin de plus de 24 à plus de 400 nucléotides, selon les différentes lignées cellulaires (Wright et al. 1997; Makarov et al. 1997; McElligott et Wellinger 1997; Huffman et al. 2000; Cimino-Reale et al. 2001). La méthode de T-OLA modifiée a été essayée sur des substrats télomériques de plus de 60 nucléotides afin de vérifier son efficacité et sa spécificité. Des fragments d'ADN contenant des séquences télomériques de différentes tailles ont été purifiés à partir des plasmides. Ces fragments contenant les répétitions télomériques sont dénaturés par la chaleur avant de procéder aux réactions de T-OLA modifiée. Les résultats obtenus (données non-montrées) concordent avec les données publiées par un autre groupe (Stewart et al. 2003) suite à l'utilisation de ces plasmides. Les résultats permettent de conclure que l'efficacité de la réaction de T-OLA modifiée permet la récupération de longs produits de ligation. Ces conditions devraient aussi permettre la récupération de possibles longs produits de ligation, lorsque la réaction sera faite sur l'ADN télomérique

issu de lignées cellulaires humaines. Par contre, l'efficacité de récupération des longs produits est diminuée par rapport aux produits plus courts.

Les réactions de T-OLA ont lieu à l'extrémité 3' simple-brin de l'ADN purifié de cellules HeLa S3

Les réactions de T-OLA modifiées ont été essayées sur l'ADN purifié de cellules HeLa afin de vérifier si la réaction a lieu à l'extrémité 3' simple-brin de l'ADN télomérique purifié. Comme contrôle négatif, l'ADN purifié a été traité avec des exonucléases spécifiques (Mung Bean Nuclease, qui dégrade spécifiquement les extensions simple-brin de l'ADN et exonucléase I de *E. coli* qui catalyse l'enlèvement de nucléotides sur un substrat d'ADN simple-brin de 3' vers 5') afin d'avoir de l'ADN sans extensions télomériques. Cet ADN traité par les exonucléases ne devrait plus permettre l'hybridation du ligand C, et par conséquent, aucune réaction de T-OLA ne devrait être possible sur cet ADN. La figure 29 montre que la récupération de produits de ligation est possible seulement lorsque l'extension 3' simple-brin est intacte.

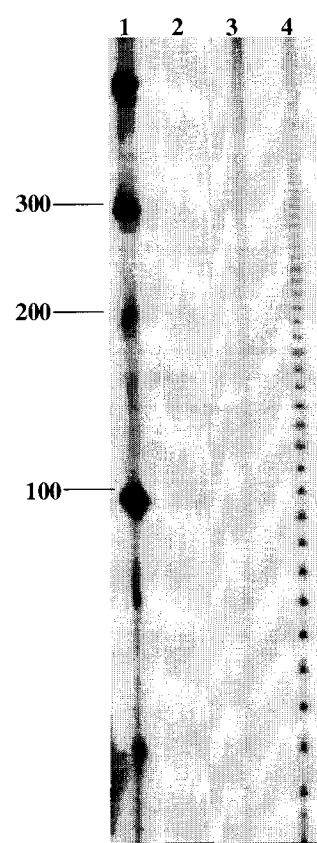


Figure 29. Les réactions de T-OLA ont lieu sur la portion 3' simple-brin des extrémités télomériques de l'ADN purifié de cellules HeLa S3. Piste 1 : marqueur de poids moléculaire 100 pb radiomarqué. Piste 2 : 5 µg d'ADN génomique purifié de cellules HeLa S3, traité avec la Mung Bean Nucléase. Piste 3 : 5 µg d'ADN génomique purifié de cellules HeLa S3, traité avec l'exonucléase I de *E. coli*. Piste 4 : 5 µg d'ADN génomique purifié de cellules HeLa S3.

La taille des extensions télomériques est variable selon les différentes lignées cellulaires étudiées

Les différentes données concernant la taille des extensions simple-brin dans les cellules humaines sont variables, allant de très courtes à plus de 400 nucléotides (Makarov et al. 1997; Wright et al. 1997; McElligott et Wellinger 1997; Huffman et al. 2000; Cimino-Reale et al. 2001). La présence d'une extension adéquate est déterminante pour la protection des chromosomes (Blackburn 2000) et est également importante en ce qui a trait à la sénescence répllicative. La méthode de T-OLA modifiée pourrait permettre une plus grande précision dans la détermination de la taille de ces extensions, et permettre de mieux comprendre la dynamique présente aux extrémités des chromosomes. De plus, certaines lignées cellulaires (ALT) sont capables de maintenir la taille de leurs télomères sans avoir recours à la télomérase (Hwang et Kucherlapati 1983; Bryan and Reddel 1997; revue dans Reddel et al. 1997). Aucune donnée concernant la dynamique des extensions simple-brin présentes aux télomères de ces cellules n'est connue à ce jour.

Étant donné que les réactions T-OLA modifiées pouvaient nous indiquer la taille des extensions simple-brin chez les cellules HeLa S3, des essais ont été entamés sur l'ADN purifié à partir de différentes lignées de cellules humaines (tableau 7). Les réactions de T-OLA sur l'ADN cellulaire purifié ont été mises sur gel 5 % polyacrylamide dénaturant, et le gel a été exposé sur autoradiogramme pour 7 jours. La taille des produits de ligation obtenus sur les extensions simple-brin peut être estimée en les comparant avec les tailles des produits de ligation obtenus avec le substrat modèle TS10 (figure 30). Pour les produits de ligation obtenus avec le substrat TS10, des bandes supplémentaires à celles récupérées jusqu'à maintenant sont observées sur

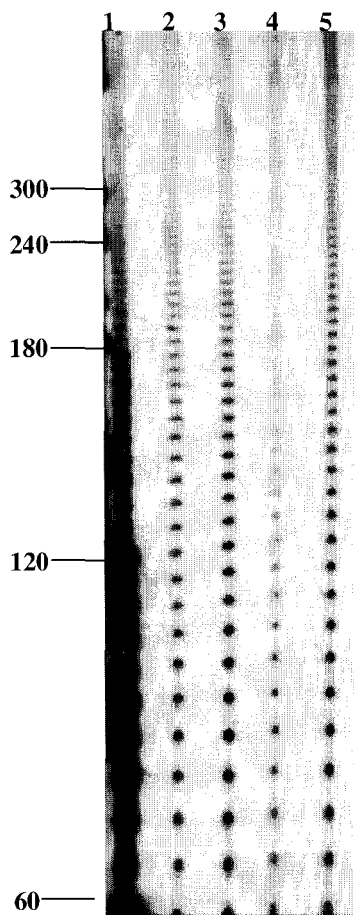


Figure 30. Réactions de T-OLA sur l'ADN purifié de différentes lignées cellulaires (1). Piste 1 : substrat modèle TS10. Piste 2 : ~5 µg d'ADN issu de la lignée ECR 293. Piste 3 : ~5 µg d'ADN issu de la lignée HeLa S3. Piste 4 : ~5 µg d'ADN issu de la lignée HCT 116. Piste 5 : ~5 µg d'ADN issu de la lignée BJ, à un passage précoce.

l'autoradiogramme exposé pour 7 jours. Ces bandes supplémentaires diminuent en intensité avec leur augmentation en taille, et elles sont les multiples parfaits de 60 nucléotides (figure 30, piste 1). Ces bandes peuvent être dues à des réactions de ligation qui n'ont jamais été détectées auparavant car elles ne sont pas très efficaces et l'exposition de courte durée (16-24 heures) ne permettait pas de les détecter. Elles peuvent être expliquées par la ligation d'un oligo ligand C hybridé à deux molécules de TS10; l'oligo ligand C phosphorylé peut servir de pont entre les deux molécules TS10 et permettre la ligation continue des oligos sur deux molécules de substrat juxtaposées. Il peut aussi y avoir de la dégradation des différents oligos menant à de plus courtes molécules. Ces molécules peuvent être hybridées et permettre la formation de « ponts » entre les molécules via le ligand C, et ainsi donner des produits de ligation qui sont de taille supérieure au substrat initial.

La figure 30 montre un résultat typique obtenu par T-OLA, soit les produits de ligation de différentes tailles, avec une différence de 6 nucléotides entre chacun (longueur d'un oligo ligand C). La taille maximale et l'intensité relative des différents produits de ligation récupérés diffèrent peu selon les lignées cellulaires utilisées ici, avec la migration et la résolution des produits permises par le gel. Pour la lignée cellulaire ECR 293, la taille

maximale atteinte par des produits de ligation se situe au-dessus de 240 nucléotides, avec la majorité des produits de ligation ayant une taille variant entre 60 et 130 nucléotides (figure 30, piste 2). Pour la lignée cellulaire HeLa S3, la taille maximale atteinte par des produits de ligation se situe également au-dessus de 240 nucléotides. L'intensité des bandes de produits entre 60 et 150 nucléotides semble indiquer une majorité de produits situés autour de ces tailles (figure 30, piste 3). Pour la lignée cellulaire HCT 116, les produits de ligation récupérés ont une faible intensité. Ceci pourrait s'expliquer par de plus courtes extensions simple-brin présentes aux télomères de cette lignée, par des réactions de T-OLA moins efficaces sur cet ADN, ou encore par une concentration d'ADN surestimée pour cette lignée lors des réactions. La taille maximale des produits de ligation obtenus suite à la réaction de T-OLA se situe autour de 150 nucléotides; de ceux-ci, la majorité se situe entre 60 et 100 nucléotides (figure 30, piste 4). Pour la lignée cellulaire BJ, l'intensité des bandes correspondant aux produits de ligation est relativement plus forte que pour les autres lignées cellulaires. La taille maximale des produits de ligation récupérés est également au-dessus de 240 nucléotides et la majorité de ces produits se situe entre 60 à 150 nucléotides (figure 30, piste 5).

La figure 30 montre une distribution d'intensité de bandes très similaire à celle qui a été décrite auparavant (Cimino-Reale et al. 2001). Dans cette étude, les bandes les plus intenses sont situées de 60 à 120 nucléotides, et l'intensité des bandes situées au-delà de 120 nucléotides semble diminuée en comparaison avec les autres bandes. L'intensité de chacune des bandes obtenues a été dosée et normalisée par rapport à la somme des intensités de toutes les bandes des produits de ligation. L'interprétation de ces données suggère que plus une bande est intense, plus il y a de produits de ligation de cette

longueur, représentant la taille des extensions. Cependant, pour que cette observation soit vraie, il faut présumer que toutes les réactions impliquées dans le T-OLA ont été accomplies avec une efficacité de 100% (i.e. marquage des oligos, hybridation sur le substrat et ligation des oligos). Comme il a été mentionné auparavant, certaines conditions peuvent favoriser l'accumulation de courts produits de ligation et biaiser l'interprétation des résultats. Les données préliminaires de mesure du simple-brin obtenues par le T-OLA modifié ne permettent pas de confirmer que les produits de 120 nucléotides et moins sont majoritaires dans la population. Certaines conditions favorisant l'accumulation de courts produits ont été améliorées. Le rapport de l'intensité de chacune des bandes obtenues par rapport à l'intensité totale de toutes les bandes dans le patron d'échelle est le résultat qui donnerait l'information sur la distribution des différentes tailles des extensions. Cependant, il reste impossible de déterminer hors de tout doute si la taille de la majorité des extensions est inférieure à 120 nucléotides, ou si la réaction de T-OLA est seulement plus efficace pour des produits plus courts que longs.

Pour vérifier la taille maximale des extensions télomériques, les produits d'une même réaction de T-OLA ont été chargés sur le gel en 2 échantillons séparés dans 2 piste juxtaposées, à environ 2 heures d'intervalle. Cette procédure permet de visualiser la taille des plus longs produits suite à la migration de la réaction pour une longue période, et aussi de visualiser les courts produits suite à la migration des échantillons chargés en 2^e sur le gel. Par contre, le fait de séparer une même réaction de T-OLA en deux échantillons à charger sur le gel diminue l'intensité des signaux finaux récupérés pour chaque migration. Les différents contrôles effectués et le marqueur de poids moléculaire sont aussi chargés sur le même gel en 2 différents échantillons. La figure 31 confirme les

données obtenues auparavant concernant la taille des extensions simple-brin de l'ADN issu des lignées cellulaires BJ et HCT 116, et permet de visualiser une taille des extensions de plus de 200 nucléotides pour chacune de ces lignées cellulaires. Elle montre également la taille des extensions télomériques simple-brin pour la lignée GM 847, une lignée de cellules ALT. La taille des extensions simple-brin détectées suite au T-OLA atteint un maximum d'environ 100-120 nucléotides. Ces données constituent une première estimation de la taille des extensions simple-brin pour des cellules immortalisées maintenant la taille des télomères par un mécanisme alternatif à l'activation de la télomérase. Les extensions simple-brin de ces cellules semblent plus courtes que celles des cellules immortalisées ayant une télomérase active ou celles de cellules en division. Étant donné que la télomérase, enzyme qui rallonge habituellement les extensions simple-brin, est inactive dans cette lignée, la génération des extensions peut être uniquement obtenue par la dégradation du brin 5' C-riche et/ou recombinaison homologue sur différents substrats.

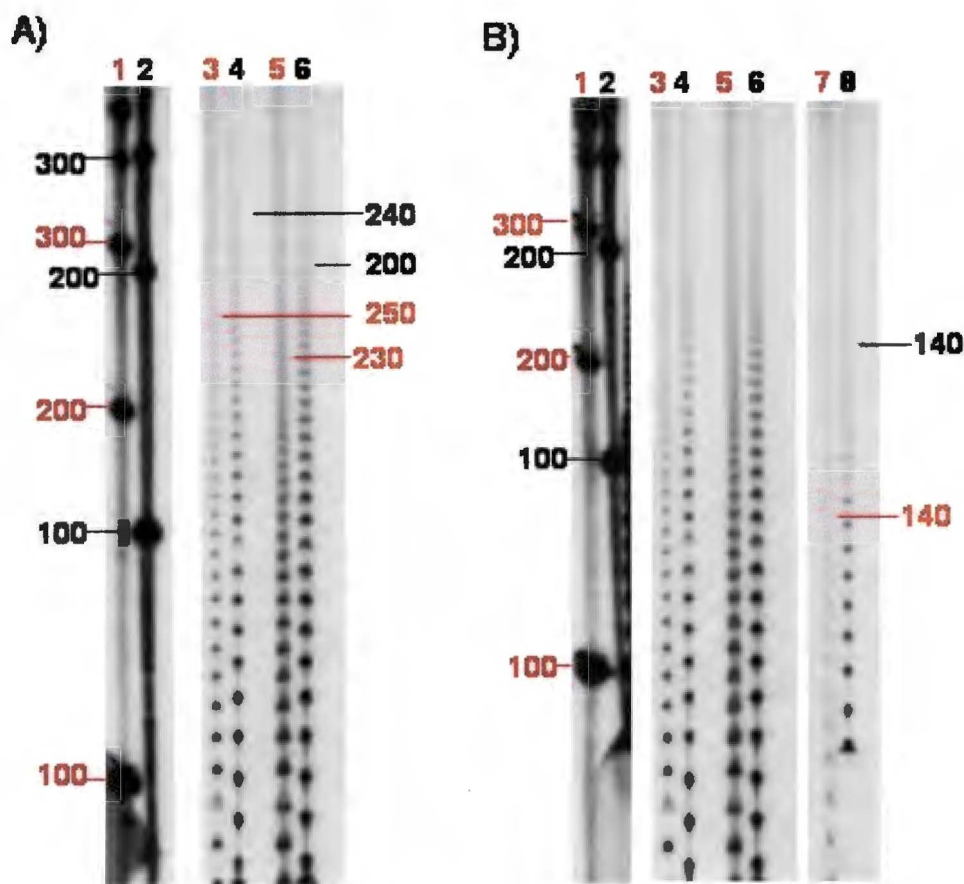


Figure 31. Réactions de T-OLA sur l'ADN purifié de différentes lignées cellulaires (2). Les réactions de T-OLA pour chacun des échantillons ont été divisées en deux. Un premier aliquot a été chargé sur le gel et migré pour 2 heures (pistes identifiées en rouge). Le second aliquot a été chargé sur le gel après les 2 heures de migration de l'échantillon précédent, puis migré encore 2 heures (pistes identifiées en noir). Les tailles des contrôles de poids moléculaire sont indiquées à gauche et les tailles des produits de ligation sont indiquées à droite (rouge= migration longue, noir= migration courte). A) Piste 1 : marqueur de poids moléculaire 100 pb radiomarqué. Piste 2 : marqueur de poids moléculaire 100 pb radiomarqué. Piste 3 : ~5 µg d'ADN issu de la lignée BJ. Piste 4 : ~5 µg d'ADN issu de la lignée BJ. Piste 5 : ~5 µg d'ADN issu de la lignée HCT 116. Piste 6 : ~5 µg d'ADN issu de la lignée HCT 116. B) Piste 1 : marqueur de poids moléculaire 100 pb radiomarqué. Piste 2 : marqueur de poids moléculaire 100 pb radiomarqué. Piste 3 : ~5 µg d'ADN issu de la lignée BJ. Piste 4 : ~5 µg d'ADN issu de la lignée BJ. Piste 5 : ~5 µg d'ADN issu de la lignée HCT 116. Piste 7 : ~5 µg d'ADN issu de la lignée GM 847. Piste 8 : ~5 µg d'ADN issu de la lignée GM 847. Les pistes 3 à 6 de la figure B sont les mêmes que celles de la figure A, et le temps d'exposition est le même pour les pistes 3 à 8.

DISCUSSION

Interactions entre les molécules hnRNP A1 et implications possibles aux télomères

La protéine hnRNP A1 lie l'ARN et a des fonctions importantes dans au moins deux mécanismes cellulaires différents, qui sont la modulation de l'épissage alternatif (Mayeda et Krainer 1992; Caceres et al. 1994; Yang et al. 1994; Blanchette et Chabot 1999; Hutchison et al. 2002; Nasim et al. 2002) et la biogénèse des télomères (LaBranche et al. 1998). Le domaine N-terminal de la protéine est responsable de la liaison spécifique à l'ARN (UP1) (Dreyfuss et al. 1993) et possède aussi les fonctions de la protéine complète quant à l'action aux télomères (LaBranche et al. 1998; Fiset et Chabot 2001; Dallaire et al. 2000; revue dans Ford et al. 2002). Étant donné que la protéine A1 est impliquée dans deux processus cellulaires différents, notre objectif dans un premier temps était de trouver des partenaires d'action de hnRNP A1/UP1 afin de mieux comprendre son mode d'action.

Au début de mes travaux, aucune donnée sur les interactions de hnRNP A1/UP1 avec des protéines impliquées aux télomères n'était connue. En effet, à ce moment, il était connu que la protéine hnRNP A1 ainsi que la portion UP1 sont capables de promouvoir l'allongement des télomères dans une lignée cellulaire dépourvue de protéine A1 (LaBranche et al. 1998). Aussi, tout comme la protéine complète, la portion UP1 peut lier les répétitions télomériques humaines *in vitro* (LaBranche et al. 1998). Par contre, uniquement la portion UP1 permet de détecter l'activité télomérase mesurée par essai TRAP, à partir d'un lysat cellulaire (LaBranche et al. 1998). Même s'il est peu probable que la protéine UP1 *per se* existe dans les cellules humaines *in vivo*, au moins en ce qui concerne l'implication de hnRNP A1 aux télomères, cette portion possède les fonctions

de la protéine complète et peut servir de modèle afin d'étudier l'implication de la protéine dans la modulation de la longueur des télomères.

Nous avons utilisé le système double-hybride chez la levure (Fields et Song 1989; Chien et al. 1991; Grossel et al. 1999) qui est fréquemment utilisé comme première approche afin de trouver des partenaires cellulaires à une protéine d'intérêt. Les interactions entre protéines détectées par cette approche doivent par la suite être confirmées par d'autres méthodes d'analyse d'interactions et/ou des analyses biochimiques. Notre approche fut d'utiliser une protéine de fusion avec la portion UP1 et de cribler une librairie d'ADNc de cellules HeLa afin de récupérer des protéines qui interagissent avec hnRNP A1/UP1. Suite aux différentes étapes requises pour la mise en place du système et le déroulement du criblage, 68 clones finaux ont été récupérés. Tous ces clones ont montré une interaction entre protéines qui était reproductible et qui dépendait de la présence de la protéine UP1. Le nombre final de clones repêchés par cette approche représente seulement 0.00027% du nombre total de clones criblés. Les plasmides récupérés ont été séquencés partiellement afin de permettre l'identification de la protéine encodée (tableau 4).

Il se peut que certaines des protéines repêchées par notre criblage soient des faux positifs concernant leur interaction avec hnRNP A1/UP1 (Bartel et al. 1993). Plusieurs protéines sont connues pour être récupérées lors de criblages sans avoir d'interaction directe avec la protéine d'intérêt, mais aucune des protéines récupérées par notre criblage n'est présente sur ces listes. Cependant, cette donnée ne veut pas dire que toutes les interactions récupérées par le criblage sont des interactions réelles qui ont lieu dans la cellule. Comme certaines protéines ont été récupérées plus souvent que d'autres lors du

criblage (tableau 4) et il est facile de croire que les interactions qui ont été détectées entre hnRNP A1/UP1 et ces différentes protéines sont réelles. Il se peut que, durant l'étape d'amplification des clones indépendants contenus dans la librairie, certains clones qui sont faiblement présents ne soient pas amplifiés, et que ceux-ci soient perdus par la suite. Également, il est possible que lors de cette étape, les protéines abondantes soient amplifiées de façon exagérée, et donc qu'il y ait un biais dès le départ dans cette méthode. Aussi lors de la purification de l'ADN amplifié à partir de la librairie, il peut y avoir perte de certains clones peu abondants. Aucune des douze différentes protéines repêchées et identifiées suite à ce criblage n'est une protéine ayant une implication spécifique connue aux télomères. Cependant, les clones récupérés sont nouveaux, et leur interaction avec hnRNP A1/UP1 peut indiquer une nouvelle implication de ces protéines, soit au niveau d'une action possible aux télomères ou encore dans l'épissage alternatif. De nombreuses expériences additionnelles dans des cellules de mammifères pourraient être envisagées afin de mieux caractériser ces interactions et leurs implications dans la cellule.

Notre criblage a permis de détecter une interaction entre les protéines hnRNP A1 et /ou UP1 elles-mêmes. Dans un article du groupe de Biamonti publié en 1996, le système double-hybride est également utilisé afin de vérifier les interactions entre des protéines hnRNPs et certains de leurs domaines (Cartegni et al. 1996). Avec ce système, les interactions entre hnRNP A1 et UP1 ne semblent pas efficaces (Cartegni et al. 1996). Cependant, les plasmides utilisés par ce groupe pour la production des protéines de fusion sont différents de ceux que nous avons utilisés, et la quantité de protéine de fusion produite peut varier grandement selon les plasmides et influencer par la suite la quantité

de protéines produites et la détection des interactions. Les souches de levure utilisées pour la détection des interactions sont aussi différentes dans les deux systèmes. Par conséquent, les promoteurs en amont des gènes rapporteurs sont différents, ce qui pourrait expliquer les différences entre les résultats obtenus dans nos systèmes respectifs. De plus, suite à plusieurs contrôles effectués dans notre système, nous avons décidé d'utiliser comme seuil critique un bas niveau d'activation des gènes rapporteurs pour notre criblage, ce qui permet la détection de plus faibles interactions. Nous aurions pu également tenter d'utiliser une procédure adaptée pour la détection des interactions faibles ou transitoires, qui augmente les chances de détecter ces interactions. Des essais d'interaction directe entre différents domaines de hnRNP A1 en utilisant notre système ont montré que l'interaction entre hnRNP A1 et UP1 est faible (figure 21), ce qui est en accord avec les données précédentes (Cartegni et al. 1996). Ces essais ont également permis de montrer une forte interaction des protéines A1 entre elles, ce qui a confirmé les résultats précédents (Cartegni et al. 1996), et a permis la mise en place d'une autre méthode pour l'analyse des domaines de la protéine A1.

Il demeure intrigant que des protéines de la famille des hnRNPs aient des fonctions particulières aux télomères. En effet, il a été montré que les protéines hnRNP A1, A2/B1, D et E et les protéines homologues aux hnRNPs chez d'autres organismes peuvent s'associer à des oligonucléotides contenant des répétitions télomériques humaines simple-brin (Dallaire et al. 2000; Ding et al. 1999; Erlitzki et Fry 1997; Ishikawa et al. 1993; Johnston et al. 1999; McKay and Cooke 1992a; Sarig et al. 1997). Les données expérimentales montrent que certaines hnRNPs peuvent lier les extensions télomériques G-riches *in vitro*, mais les données concernant cette liaison *in vivo* restent à

établir. Comme les protéines hnRNPs sont nucléaires et abondantes dans la cellule, il est possible qu'elles aient accès à l'ADN télomérique. La protéine hnRNP A1/UP1 peut protéger l'extension simple-brin de la dégradation (Dallaire et al. 2000), tandis que hnRNP D peut, en plus de sa liaison au simple-brin, déstabiliser des pairages entre les guanines de la séquence télomérique (G-G pairing) (Eversole et Maizels 2000). Cette observation permet de suggérer que hnRNP D pourrait être responsable de l'enlèvement de structures potentielles formées par l'extension simple-brin riche en G, et faciliter l'allongement par la télomérase lors de la réplication des télomères. À ce sujet, il a été montré que hnRNP A1/UP1, hnRNP D, hnRNP C1/C2 et la protéine La peuvent copurifier avec l'activité télomérase (LaBranche et al. 1998; Eversole et Maizels 2000; Ford et al. 2000; Ford et al. 2001; Ford et al. 2002). Les résultats de ces études laissent croire que les protéines hnRNPs auraient un rôle dans la stimulation et/ou la répression de la télomérase (Dallaire et al. 2000; Ford et al. 2001; LaBranche et al. 1998). Il a d'ailleurs été montré que hnRNP A1 peut lier l'ARN de la télomérase et les extensions G-riches (Fiset et Chabot 2001), ce qui peut suggérer un rôle semblable à celui de Est1p chez la levure (Evans et Lundblad 1999; Evans et Lundblad 2002; Steiner et al. 1996; Zhou et al. 2000), pour le recrutement de la télomérase à son site d'action sur le télomère. Vu que les hnRNPs sont souvent retrouvées en complexes (LeStourgeon et al. 1978), il est possible qu'un groupe de hnRNPs lie la télomérase et que d'autres hnRNPs s'associent à la télomérase et aux télomères. Cependant, dans les cellules humaines, la protéine hPot1p lie le même substrat que les hnRNPs, soit le simple-brin télomérique (Baumann et Cech 2001; Loayza et de Lange 2003; Colgin et al. 2003) et elle peut aussi être un régulateur de l'activité de la télomérase (Armbruster et al. 2004; Veldman et al. 2004; Kelleher et al.

2005). Étant donné que les hnRNPs lient l'ARN avec une plus grande affinité que l'ADN grâce aux domaines RRM présents dans ces protéines, il se peut que leur liaison aux télomères soit le reflet d'un manque de spécificité dans la liaison. La longueur des télomères est clairement affectée par des changements dans le niveau des protéines hnRNP A1 et La, mais il demeure possible que cet effet soit indirect. En effet, des mutations ou des délétions de ces protéines peuvent affecter l'expression de plusieurs différents gènes de multiples façons, ce qui peut mener à des effets secondaires indirects sur la télomérase et les télomères. La délétion des protéines A1 et A2 dans des lignées de cellules transformées permet de promouvoir l'apoptose, ce qui n'est pas le cas dans des cellules normales (Patry et al. 2003). La même observation peut être faite dans des lignées de cellules de souris (Patry et al. 2004), mais ce processus pourrait être causé par la délétion des protéines comme telles. Avec notre approche de criblage utilisant hnRNP A1/UP1 comme appât, aucune protéine ayant une implication directe aux télomères n'a été récupérée, ce qui mène à questionner l'action directe de hnRNP A1/UP1 dans la biogénèse des télomères. Un modèle basé sur les données expérimentales de l'action des hnRNPs aux télomères a été élaboré et il permet de réconcilier les données concernant les hnRNPs et leur action possible aux télomères en deux modèles (figure 32). D'autres

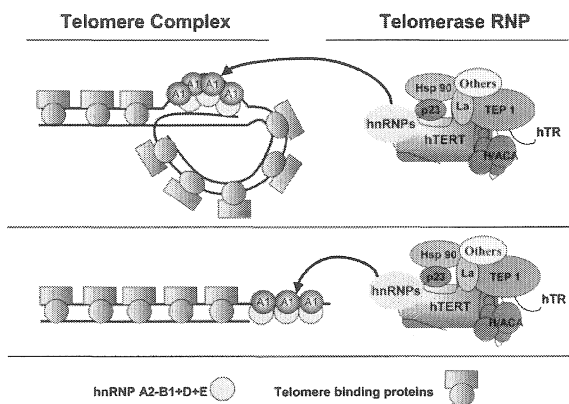


Figure 32. Modèle d'action des protéines hnRNPs aux télomères. L'holoenzyme télomérase est montré à droite avec hTR, hTERT et les autres protéines associées. Le télomère est montré dans la forme t-loop (haut) où les hnRNPs sont associées à la portion simple-brin déplacée. La forme linéaire du télomère (bas) montre les hnRNPs liées sur l'extension. La flèche entre les hnRNPs présentes dans l'holoenzyme télomérase et les hnRNPs présentes sur le télomère indique que des interactions peuvent avoir lieu, menant à la formation de complexes et possiblement au recrutement de la télomérase au télomère. Tiré de Ford et al. (2002)

évidences sont cependant requises avant de confirmer ce modèle d'action.

Interactions de hnRNP A1 et modèle d'action dans l'épissage alternatif

La protéine hnRNP A1 possède la capacité de moduler l'épissage alternatif (Mayeda et Krainer 1992; Caceres et al. 1994; Yang et al. 1994; Blanchette et Chabot 1999; Hutchison et al. 2002; Nasim et al. 2002). Un modèle expliquant la modulation de l'épissage alternatif par la liaison de A1 à l'ARN pré-messager et l'interaction entre les protéines a été proposé en prenant comme transcrit modèle l'ARN pré-messager de A1 lui-même (Blanchette et Chabot 1999) (figure 18). Il stipule que la liaison de A1 aux sites optimaux CE1a et CE4 (figure 17), situés de part et d'autre de l'exon alternatif, serait suivie d'une interaction protéine : protéine impliquant le domaine GRD de A1. Cette interaction entre les protéines liées à l'ARN provoquerait le rapprochement dans l'espace des partenaires d'épissage distants, menant à la sélection du site d'épissage distal. En accord avec ce modèle, l'épissage *in vitro* d'un ARN dans lequel les sites de liaison pour A1 ont été remplacés par des séquences formant une structure duplexe permet aussi la sélection du site d'épissage distal (Nasim et al. 2002). Étant donné que la formation de duplexe ne survient pas de façon efficace chez les mammifères (Solnick et Lee 1987; Eperon et al. 1988), certaines protéines dont A1 pourraient avoir comme tâche de moduler la conformation des ARNs pré-messagers pour assurer leur épissage (Blanchette et Chabot 1999). En ce sens, il a été montré que A1 a la capacité de se lier à un ARN fixé à une matrice tout en étant capable de lier un second site de liaison (Blanchette et Chabot 1999). Cette interaction requiert la présence du domaine GRD car UP1 peut se lier à un site de liaison optimal fixé à la matrice mais ne peut maintenir un second ARN libre

(Blanchette et Chabot 1999), pouvant suggérer possiblement l'implication d'un dimère de protéines.

Notre criblage utilisant un système double-hybride a permis de mettre en évidence une interaction forte entre au moins deux protéines hnRNP A1 (figure 21), tel que montré auparavant par Cartegni et al. (1996). Ce système double-hybride a servi afin de caractériser les domaines protéiques impliqués dans l'interaction entre les protéines. Il a également été utilisé afin de tester si cette interaction a un rôle à jouer dans la modulation de la sélection des sites d'épissage effectuée par A1. Diverses approches de troncations de protéines, de mutagénèses dirigée et aléatoire ont été élaborées pour analyser l'interaction entre les protéines A1 et des domaines requis pour celle-ci. La figure 21 et les tableaux 5 et 6 résument les différentes données générées suite à ces approches.

Avec la première approche utilisée, les différents domaines de la protéine A1 ont été analysés pour leurs interactions. Les résultats indiquent qu'avec le système double-hybride utilisé, les protéines A1 complètes interagissent fortement entre elles, et que cette interaction est dépendante de la présence du domaine GRD. En effet, les portions UP1 ont une interaction entre elles qui est faible, tandis que les portions GRD peuvent interagir entre elles, mais cette interaction est environ 5 fois plus faible que celle entre les protéines A1 complètes. Ceci permet également de suggérer que la présence du domaine UP1 est importante pour permettre des interactions fortes entre les protéines. Cette portion UP1 est importante pour la liaison à l'ARN et le mécanisme d'action de la protéine complète dans l'épissage. En effet, même si le domaine GRD peut interagir avec lui-même, il a été montré que le recrutement artificiel du domaine GRD à des sites optimaux situés de part et d'autre d'un exon alternatif, n'est pas en mesure de moduler la

sélection des sites d'épissage dans le système utilisé (G. Lettre, M. Blanchette et B. Chabot, données non-publiées). De plus, le recrutement de la portion UP1 seule à des sites de liaison optimaux de part et d'autre d'un exon alternatif est également incapable de moduler la sélection des sites d'épissage (Mayeda et al. 1994; Blanchette et Chabot 1999). La portion UP1 pourrait donc agir en permettant de stimuler ou de stabiliser l'interaction entre les protéines. Les domaines GRD des protéines A1 seraient responsables d'initier les interactions entre les protéines A1 liées à l'ARN, mais ils possèderaient également des fonctions particulières pour permettre la modulation de l'épissage. En effet, si le domaine GRD est remplacé par le domaine RS du facteur d'épissage ASF/SF2 dans une protéine A1 chimère (UP1 + RS de ASF/SF2), les protéines résultantes perdent leur capacité à moduler l'épissage alternatif (Mayeda et al. 1994). Une fois les interactions entre protéines initiées, des interactions additionnelles entre les portions UP1 de A1 pourraient contribuer à stabiliser les protéines. Il a été montré auparavant que chacun des deux RRM de A1 est requis pour sa fonction dans l'épissage alternatif : chacun de ces RRM joue un rôle distinct dans ce processus et leur position relative est importante pour le mécanisme (Mayeda et al. 1998). Malgré le fait que les 2 RRM soient presque similaires et symétriques, leurs fonctions ne sont pas équivalentes (Mayeda et al. 1998).

Avec la deuxième approche (mutations dirigées), les différents résidus ciblés étaient répartis en trois classes, basées sur leur positionnement dans le cristal d'UP1 avec un oligo télomérique (figure 22) (Ding et al. 1999). Une première classe était celle de résidus qui semblaient être importants pour l'interaction entre les molécules UP1, une deuxième classe consistait en un résidu important dans la liaison des acides nucléiques et

une troisième classe consistait en un résidu quelconque qui n'est pas impliqué dans l'établissement d'interactions. Une approche semblable avait été entreprise auparavant et avait permis le remplacement des résidus phénylalanines présents dans les RRM_s (Mayeda et al. 1994). Les variations dans la capacité des protéines à lier l'ARN étaient minimales suite à ces mutations, mais toutes les protéines mutantes analysées montraient une perte totale de leur capacité à moduler l'épissage alternatif (Mayeda et al. 1994). Dans nos études, les différentes mutations introduites dans les protéines A1 et même la combinaison de plusieurs de ces mutations n'ont eu aucun effet majeur sur les fonctions de A1 qui ont été étudiées (tableau 5). Même si les résidus visés par les mutations sont importants pour les interactions entre les monomères d'UP1 ou le contact avec l'ADN simple-brin dans le cristal (Ding et al. 1999), les protéines recombinantes produites et testées se comportent comme les protéines A1 de type sauvage. À la lumière d'une autre étude sur les RRM_s de UP1 et leur implication dans la liaison des acides nucléiques (Myers et Shamoo 2004), d'autres positions de choix (Lys 106, Phe 108, Glu 135, Phe 150, Arg 178, Lys 179, Leu 181, Lys 183) auraient pu subir des mutations afin de mieux comprendre par la suite leur implication dans la sélection des sites permettant la modulation de l'épissage.

Avec la troisième approche (mutations aléatoires), les mutations introduites dans la protéine A1 étaient laissées au hasard. Suite à la procédure établie, environ 2000 colonies contenant un fragment d'ADN_c mutagénisé encodant la protéine A1 ont été obtenues. Dans un premier temps, les protéines de fusion mutantes obtenues dans la levure ont été analysées pour leur capacité à interagir avec une protéine GAL4-AD-A1 de type sauvage contenue dans la souche de levure. Des 2000 colonies de départ, 244

colonies ayant perdu cette capacité ont été récupérées. De ces colonies, 28 contenaient des protéines de fusion GAL4 DBD-A1 mutantes complètes. Certains des ADNc encodant ces protéines ont été clonés dans le vecteur d'expression pGEX-KG et les protéines GST-A1 mutantes produites ont été analysées pour leur capacité à interagir avec un ARN contenant un site de haute affinité pour A1. Celles qui interagissaient encore avec l'ARN malgré les mutations ont été analysées pour leur capacité à moduler l'épissage d'un transcrit modèle. Le tableau 6 dresse un résumé des différentes mutations obtenues dans la protéine A1, et des résultats obtenus suite aux analyses faites avec ces protéines.

Étant donné l'importance du domaine GRD pour permettre l'interaction entre protéines, il aurait été attendu de récupérer des protéines A1 mutantes contenant des mutations dans le domaine GRD causant une perte d'interaction. La proportion de glycines dans ce domaine est de 50, sur un total de 125 résidus (40%). La procédure de mutagenèse utilisée ne semble pas permettre de générer assez de mutations dans ce domaine pour abolir l'interaction. La présence d'une haute proportion de glycine dans le GRD empêche de voir l'implication de chaque résidu dans l'interaction étant donné leur présence et la possible redondance de leur fonction. Il est cependant possible de spéculer que le domaine GRD est requis dans sa totale intégrité. En effet, le mutant 16 montre un changement du cadre de lecture en position 256 et un codon stop introduit en position 276 de la protéine A1 (total de 320 acides aminés dans A1). Ces mutations mènent à un phénotype de perte d'interaction entre les protéines. Les mutations combinées de ce mutant ne permettent pas de dire si la troncation du GRD à cet endroit mène à la perte d'interaction. En effet, le changement du cadre de lecture qui est observé élimine environ

la moitié des résidus glycines du GRD (21 sur 50) et l'introduction du codon stop mène à la perte de 15 résidus glycines.

Les mutations aléatoires obtenues dans la séquence de A1 qui ont permis de récupérer des protéines complètes peuvent être divisées en deux classes. Premièrement, les mutations qui se retrouvent dans le RRM1 et deuxièmement, les mutations qui se retrouvent dans le RRM2. Les RRM de A1 comprennent environ 90 résidus et sont faits de 4 feuillets plissés β et deux hélices α (Ding et al. 1999; Myers et Shamoo 2004). Les feuillets β anti-parallèles permettent de former la surface responsable de la liaison aux acides nucléiques (Ding et al. 1999; Myers et Shamoo 2004). En utilisant des modifications de bases dans un oligo télomérique, J.C. Myers et Y. Shamoo ont pu sonder la surface des RRM de UP1 afin d'identifier les résidus qui permettent la reconnaissance des purines contenues dans les acides nucléiques (Myers et Shamoo 2004). Plusieurs des contacts spécifiques établis entre les bases des RRM et de l'acide nucléique sont du feuillet β 4 et des résidus en C-terminal (RRM2 et résidus en C-terminal du RRM2) (Myers et Shamoo 2004), comme identifié auparavant par le groupe de Xu (Ding et al. 1999). Les résidus très conservés présents dans les RRM et formant les RNPs se retrouvent dans les feuillets β 1 et β 3, respectivement (Ding et al. 1999; Myers et Shamoo 2004). Deux des résidus les plus grandement conservés dans les séquences RNPs sont les résidus aromatiques (Phe ou Tyr) exposés au solvant, qui permettent des interactions avec les bases de l'acide nucléique, ADN ou ARN (Myers et Shamoo 2004). Ces résidus hautement conservés sont trouvés dans toutes les structures de RRM connues et leur interaction est semblable, ce qui suggère un mécanisme

commun et important pour la reconnaissance des acides nucléiques et de leur liaison (Myers et Shamoo 2004).

Les mutations aux positions 72 et 77, localisées dans le RRM1, permettent de garder la capacité de liaison à l'ARN des protéines. Une mutation à la position 72 a été récupérée à deux reprises durant le criblage. La capacité de lier l'ARN varie selon la protéine produite, ce qui peut suggérer la présence d'une autre mutation qui n'a pas été détectée dans une des deux protéines. Les mutations aux positions 21 (dernier résidu du RNP2 présent dans le RRM1) et 58 (4^e résidu du RNP1 présent dans le RRM2) mènent à une perte de la capacité de lier l'ARN. Même si ces résidus ne semblent pas impliqués dans la liaison directe de l'acide nucléique, ils sont localisés à des endroits stratégiques permettant cette fonction. La mutation de ces résidus doit forcer un réarrangement de la structure de la protéine et les résidus importants pour le contact à l'ARN doivent se retrouver dans une conformation modifiée ne permettant plus leur contact avec l'ARN. Donc, même si les contacts entre les acides nucléiques et les RRM de UP1 sont plus importants dans le RRM2 (Myers et Shamoo 2004), d'autres positions semblent également importantes à cet effet dans le RRM1. Comme les méthodes utilisées afin d'étudier les interactions entre hnRNP A1/UP1 et les acides nucléiques diffèrent grandement entre notre étude, celle du groupe de Shamoo et celle du groupe de Xu (Ding et al. 1999; Myers et Shamoo 2004) cela peut introduire un biais dans l'importance des positions étudiées et les résultats obtenus peuvent donc varier. Malgré tout, les positions identifiées sont localisées dans des régions consensus pour les 3 études.

Les mutations retrouvées dans le RRM2 sont plus difficiles à catégoriser car elles sont survenues plus souvent en paires. Une mutation à la position 151 (6^e résidu du RNP1

présent dans le RRM2) survient à deux reprises. Ce résidu est situé à côté du résidu Phe 150, qui a été identifié comme un des résidu établissant des liens avec les acides nucléiques (Ding et al. 1999; Myers et Shamoo 2004). La mutation récupérée doit changer la conformation dans l'environnement du résidu Phe, menant possiblement à la perte d'interaction avec l'ARN. Cependant, une mutation récupérée à la position 153 (dernier résidu du RNP1 présent dans les RRM2) permet de garder la capacité d'interaction avec l'ARN comme dans une protéine A1 normale. Cette mutation a été récupérée survient à deux reprises, une fois seule et une fois combinée à une mutation au résidu 170 : la combinaison des 2 mutations abolit l'interaction de la protéine avec l'ARN, ce qui n'était pas le cas de la mutation simple. Une mutation en position 165 est aussi retrouvée à plusieurs reprises, en combinaison avec d'autres mutations. Ce résidu est situé entre 3 résidus importants qui sont impliqués dans des interactions entre les monomères d'UP1 (positions 164, 166, 167) (Ding et al. 1999), ce qui peut expliquer la perte d'interaction qui est détectée lors de sa mutation. La perte de la capacité d'interaction avec l'ARN doit dans ces cas être due à la seconde mutation présente dans les différentes protéines (positions 125, 107, 177 et 193). La position 193 est hors du RRM et à la toute fin du domaine UP1 donc n'a pas une importance majeure dans la liaison de l'acide nucléique. La position 125 est localisée dans le RRM2 et sa mutation peut changer la conformation de la protéine, menant à une perte de l'interaction avec l'ARN. La mutation en position 177 introduit un codon stop, détruisant 4 positions importantes pour l'interaction avec l'ARN (Ding et al. 1999), (Myers and Shamoo 2004). La position 107 (1^{er} résidu du RNP2 situé dans le RRM2) est située entre deux résidus

impliqués dans le contact de l'acide nucléique (positions 106 et 108) (Ding et al. 1999; Myers et Shamoo 2004).

Le modèle expliquant l'action de A1 dans la sélection des sites d'épissage alternatif est basé sur deux propriétés importantes de la protéine : sa capacité de lier des séquences de haute affinité localisées dans l'ARN et sa capacité à interagir de façon homodimérique. Les protéines mutantes obtenues par notre approche ont perdu la capacité d'établir des interactions lorsqu'elles sont testées dans le système double-hybride qui est utilisé. Par conséquent, si ces protéines sont encore capables de lier l'ARN, elles ne devraient par contre pas être en mesure de moduler la sélection d'un site d'épissage distal sur un transcrit modèle. Les protéines A1 mutantes capables de lier l'ARN ont été testées dans des essais d'épissage *in vitro* (figure 26). L'épissage d'un transcrit modèle (contenant des sites optimaux pour la liaison de A1 de part et d'autre d'un exon alternatif) dans des extraits nucléaires normaux montre une majorité de produits distaux, car A1 est encore présente dans ces extraits et son action favorise l'utilisation d'un site d'épissage distal. L'épissage d'un transcrit modèle dans des extraits dépourvus de A1 dans lesquels on rajoute des protéines recombinantes A1 montre les mêmes patrons que ceux observés pour l'épissage dans un extrait normal. Par contre, le patron d'épissage du transcrit modèle dans des extraits dépourvus de A1 est différent; à cause de l'absence de A1, il y a présence d'une majorité de produits proximaux, et une très faible proportion de produits distaux. Si le modèle proposé pour l'action de A1 dans l'épissage alternatif (Blanchette et Chabot 1999) est correct, l'épissage du transcrit modèle dans des extraits dépourvus de A1 dans lesquels on rajoute une protéine recombinante A1 mutante devrait montrer un patron semblable à celui décrit pour un

extrait dépourvu en A1 (majorité de produits proximaux et très peu de produits distaux). Cependant, les résultats d'épissage obtenus avec les protéines A1 recombinantes mutantes 37, 61 et 63, montrent un patron qui est semblable à celui observé lors du rajout de protéine A1 normale, soit une majorité de produits distaux et peu de produits proximaux. Selon ces résultats, l'interaction entre les protéines A1 ne semble pas importante pour la sélection des sites d'épissage ou alors cette interaction est importante et elle a toujours lieu entre les protéines lors des réactions d'épissage. Certaines observations permettent de mieux comprendre les résultats obtenus.

Premièrement, les pertes d'interactions entre les protéines A1 détectées au départ sont observées entre une protéine de fusion GAL4-DBD-A1 comportant des mutations et une protéine de fusion GAL4 AD-A1 de type sauvage. Les mutations peuvent engendrer un changement de conformation dans la protéine mutante qui lui fait perdre sa capacité d'interaction avec une protéine de type sauvage. Il est possible que lorsque les protéines A1 mutantes sont testées dans des réactions d'épissage, celles-ci redeviennent capables d'interagir entre elles, ce qui peut alors expliquer les résultats observés. En ce sens, les résultats de perte d'interaction n'ont jamais pu être confirmés par une autre méthode biochimique. Deuxièmement, les protéines recombinantes utilisées dans les essais d'épissage contiennent le domaine GST, qui est capable de promouvoir la dimérisation de protéines par interaction avec lui-même (Niedziela-Majka et al. 1998; Nemoto et al. 1994). De plus, il a été observé que le domaine GST existe sous forme homodimérique en solution (Walker et al. 1993) et dans des cristaux (Lim et al. 1994; McTigue et al. 1995). Le domaine GST présent en amont des protéines A1 contenant des mutations pourrait initier les contacts entre les protéines lors des essais d'épissage. Ces interactions

pourraient par la suite être stabilisées, ce qui mènerait à un patron d'épissage semblable à celui obtenu lors de l'utilisation de protéines A1 normales. Les protéines de fusion auraient dû être clivées à la thrombine puis essayées dans l'épissage *in vitro*. Cependant, la production relativement faible de ces protéines mutantes n'a pas permis de poursuivre cette direction. L'ADNc des protéines mutantes récupérées lors du criblage a été cloné dans un vecteur d'expression permettant la production de protéines recombinantes avec un domaine HIS-6, puis celles-ci ont été testées dans l'épissage (B. Labrecque et B. Chabot, données non-publiées). Les protéines recombinantes mutantes qui ont été testées (HIS-A1-37 et HIS-A1-107) ont montré une faible capacité à moduler l'épissage alternatif quand elles sont comparées à la protéine hnRNP A1 de type sauvage. Les interactions entre GAL4-DBD-A1 mutante et GAL4-AD-A1 mutante ont aussi été testées (B. Labrecque et B. Chabot, données non-publiées).

D'autres données expérimentales portent à croire que le modèle proposé est valide, malgré les résultats obtenus avec les protéines GST-A1 mutantes dans l'épissage. Le modèle implique que l'établissement d'un complexe stable entre les molécules A1 a lieu quand ces molécules sont liées à l'ARN. Cette donnée n'est pas vérifiée directement dans les essais utilisés, et il est impossible de confirmer si la présence d'ARN a un effet important sur les interactions détectées. Aussi, l'insertion de séquences dans le transcrit modèle, menant à une augmentation de la distance entre les sites de liaison de A1 et le site d'épissage 5', ne compromet pas l'utilisation de ce site (Nasim et al. 2002). De plus, le modèle proposé est aussi en accord avec le cristal de UP1 avec un oligo télomérique (Ding et al. 1999). Le modèle mis en place permet aussi d'expliquer des résultats qui ont été obtenus dans d'autres systèmes. En effet, l'insertion de 2 sites de liaison pour A1 de

part et d'autre d'un exon cassette dans l'ARN pré-messager d'adénovirus permet l'épissage au site distal (Eperon et al. 2000). Plusieurs exons cassette présents dans d'autres unités d'épissage sont aussi flanqués de sites potentiels de la liaison de A1, ce qui peut contribuer à l'utilisation de sites d'épissage distaux (Blanchette et Chabot 1999; Chabot et al. 2003). La distribution théorique du motif UAGRG^A/U lié par A1 est d'environ un par 1000 nucléotides. Par contre, une région de 600 nucléotides du gène de troponine T du rat contenant l'exon alternatif 17 contient 4 de ces sites, suggérant un rôle pour la liaison de A1 dans l'épissage de l'exon alternatif (Chabot et al. 2003).

Rôles de hnRNP A2 dans la modulation de l'épissage alternatif

Les protéines hnRNP A1 et A2 possèdent une structure fortement similaire et les protéines humaines partagent 68% d'identité (figure 16) (Mayeda et al. 1994). Chacune de ces protéines possède également une variante due à l'épissage alternatif (A1^B contient l'inclusion de l'exon alternatif 7B présent dans le gène de A1, et B1 contient 12 acides aminés en N-terminal de A2) (Mayeda et al. 1994). Des essais de chromatographie d'affinité avec des molécules ADN ou ARN contenant les sites de liaison de haute affinité pour A1 permettent aussi de lier les autres membres de la famille, soient A1^B, A2 et B1 (Caputi et al. 1999; Hutchison et al. 2002). Il a été montré que l'épissage d'un transcrit modèle (contenant les sites CE1 et CE4) dans un extrait nucléaire dépourvu des protéines de la famille hnRNPs A et B, possède le même patron d'épissage suite à l'addition de protéines recombinantes A1, A2 ou A1^B (Hutchison et al. 2002). Cependant, l'influence des différents isoformes sur l'épissage d'un ARN pré-messager naturel *in vivo* reste à démontrer.

La lignée cellulaire CB3C7 (tableau 3) exprime très peu de protéine A1 endogène. Dans des extraits nucléaires de cette lignée, le patron d'épissage d'un transcrit modèle contenant les sites CE1 et CE4 flanquant un exon cassette se compare au patron obtenu dans un extrait cellulaire normal (Hutchison et al. 2002). Si l'on procède à la déplétion des protéines hnRNP A2/B1 présentes dans un extrait de cellules CB3C7, le patron d'épissage du même transcrit modèle devient alors comparable à celui obtenu dans un extrait cellulaire dépourvu des protéines hnRNPs A et B (Hutchison et al. 2002). L'addition de protéines recombinantes A1 ou A2 à un extrait CB3C7 dépourvu des protéines A et B permet de restaurer l'utilisation du site d'épissage 5' distal d'un transcrit modèle (Hutchison et al. 2002). Ces observations suggèrent une fonction redondante des différentes protéines hnRNP A1/A1^B/A2/B1 dans l'épissage d'un transcrit modèle. Cette redondance de fonction entre A1 et A2 peut expliquer l'utilisation des sites d'épissage distaux d'un transcrit modèle dans la lignée cellulaire CB3C7 (Yang et al. 1994). Étant donné la présence de la protéine A2 dans les extraits, A2 peut remplacer A1 dans la sélection des sites d'épissage et peut permettre la production de produits distaux.

Les séquences de haute affinité A2RE-1 et A2RE-2 de VIH-1, qui sont liées par hnRNP A2, ont été analysées pour leur rôle possible dans la sélection des sites d'épissage. L'implication potentielle de ces séquences dans la modulation de l'épissage a été comparée entre les transcrits contenant les séquences A2RE et les transcrits contenant des sites de liaison de haute affinité pour A1. Les transcrits contenant les séquences A2RE-1 ou A2RE-2 sont épissés majoritairement au site d'épissage 5' proximal, comme c'est le cas pour le transcrit contrôle ne contenant aucun site de liaison pour A1 (Bériault et al. 2004). Ces données suggèrent que la protéine A2 ne se lie pas à ces séquences dans

ce contexte, ou qu'elle s'y lie mais ne peut pas influencer l'épissage (Bériault et al. 2004). Ces séquences sont incapables de permettre la liaison de la protéine A1, dans ce contexte (Bériault et al. 2004). Des essais d'épissage dans des extraits nucléaires dépourvus des protéine hnRNPs A et B dans lesquels on rajoute les protéines recombinantes A1 ou A2 montrent également que les séquences A2RE insérées dans le transcrit modèle n'ont aucune influence sur la sélection des sites d'épissage (Bériault et al. 2004).

Les protéines hnRNP A1 et A2 peuvent avoir des fonctions redondantes en ce qui concerne l'épissage d'un transcrit modèle contenant des sites de liaison de haute affinité pour A1 (Mayeda et al. 1994; Hutchison et al. 2002; Chabot et al. 2003), mais des observations montrent aussi que ces protéines ont des fonctions qui leur sont propres. En particulier pour la protéine hnRNP A2 qui joue des rôles au niveau de la transcription, la maturation de l'ARN, l'épissage, le transport de l'ARN. La protéine A1 n'est pas active et ne peut pas remplacer A2 en ce qui a trait au transport de l'ARN médié par les séquences A2RE. Le transport de l'ARN de VIH-1 est dépendant de l'expression de hnRNP A2 dans les cellules, tel que montré par RNAi ciblant les protéines A1 et A2 (Sanchez-Velar et al. 2004). Également, des différences ont été montrées entre les protéines A1 et A2 en ce qui concerne l'épissage de l'ARN de SMN1 (Kashima et Manley 2003).

Régulation de l'expression de A1 et de A1^B; de nombreux éléments de contrôle

Le contrôle de l'épissage de l'ARN pré-messager de A1 résulte d'un équilibre entre plusieurs éléments de contrôle, qui agissent pour stimuler ou réprimer l'utilisation

des différents sites d'épissage pour l'incorporation de l'exon cassette 7B dans l'ARN messager produit (revue dans Chabot et al. 2003). La régulation de l'épissage de l'ARN de A1 n'est pas le seul exemple mettant en évidence un système de régulation qui implique directement les protéines produites par cet épissage. D'autres facteurs d'épissage peuvent contrôler leur propre expression en modulant la sélection de sites d'épissage, comme les facteurs Sxl et Tra de la drosophile. Dans la majorité des tissus chez l'humain, la proportion de A1 est d'environ 90% alors que celle de A1^B est de moins de 10% (Buvoli et al. 1990a; Hanamura et al. 1998). Le modèle élaboré montre que les éléments de contrôle menant à cette distribution protéique sont en fait les produits mêmes de cette boucle de régulation, i.e. que les produits de l'épissage de l'ARN de A1 vont réguler leur propre épissage. En combinant de la sorte les sites d'épissage distaux par la liaison de A1 aux éléments CE1 et CE4, la cellule produit normalement plus de A1 que de A1^B. La faible quantité de A1^B pourrait être due à des événements d'épissage minoritaires, où les éléments de contrôle sont ignorés par le spliceosome qui s'assemble. La cellule semble avoir développé des mécanismes très précis dans le but de s'assurer de la production adéquate des protéines requises. Il peut cependant y avoir certaines conditions qui demandent un changement de la proportion des protéines produites, par exemple lors du développement ou dans certains tissus particuliers. Ces protéines pourraient avoir des activités différentes qui sont requises à des moments précis. À cet effet, la cellule doit pouvoir activer d'autres éléments de contrôle lui permettant d'augmenter la production de A1^B par rapport à A1 (Simard et Chabot 2000; Simard and Chabot 2002; revue dans Chabot et al. 2003).

Existence des extensions télomériques en phase G1 du cycle cellulaire chez la levure

L'extension télomérique simple-brin a été proposée comme étant un des éléments fondamentaux permettant la formation du capuchon télomérique, et par conséquent, assurant la protection des extrémités des chromosomes (Blackburn 2000). Il a été proposé qu'une extension G-riche soit une caractéristique essentielle des chromosomes, retrouvée aux télomères de tous les eucaryotes (Henderson et Blackburn 1989; Lingner et al. 1995). De telles structures ont en effet été détectées chez les protozoaires (Klobutcher et al. 1981; Pluta et al. 1982; Henderson et Blackburn 1989; Jacob et al. 2001), l'humain (McElligott et Wellinger 1997; Makarov et al. 1997; Wright et al. 1997; Cimino-Reale et al. 2001), la souris (Hemann et Greider 1999), la plante (Riha et al. 2000) et maintenant chez la levure (Larrivée et al. 2004). Des protéines spécifiques ou des complexes de protéines spécifiques ont pour fonction la liaison de ces extensions (Gottschling et Zakian 1986; Price 1990; Sheng et al. 1995; Garvik et al. 1995; Lydall et Weinert 1995; Nugent et al. 1996; Baumann et Cech 2001; LaBranche et al. 1998), afin d'assurer la protection de l'intégrité du génome de divers organismes. La présence de ces extensions est également requise afin de permettre l'élongation par la télomérase, car *in vitro*, le substrat de la télomérase est l'ADN simple-brin (Lingner et Cech 1996).

Au début de mes travaux, les données au sujet de la présence de l'extension simple-brin chez la levure *S. cerevisiae* étaient disponibles seulement lors de la phase S du cycle cellulaire (Wellinger et al. 1993a; Wellinger et al. 1993b). L'analyse des

intermédiaires de réplication formés au niveau des télomères a montré l'existence d'extensions G-riches simple-brin d'au moins 30 nucléotides qui sont détectées tard en phase S, lors de la réplication de l'ADN télomérique (Wellinger et al. 1993a; Wellinger et al. 1993b; Marcand et al. 2000). Lors de la réplication de l'ADN par la machinerie de réplication conventionnelle, la synthèse du brin retardé créera, à une des fins du chromosome, une extension G-riche suite à l'enlèvement de l'amorce du dernier fragment d'Okazaki. Par contre, de l'autre côté du chromosome, la synthèse du brin avancé créera un bout franc (figure 6) (Lingner et al. 1995; Greider 1996; Zakian 1996). Comme il a été montré que des extensions sont présentes aux deux extrémités d'un plasmide linéaire (Wellinger et al. 1993b; Wellinger et al. 1996), il a été suggéré que le brin avancé subit une étape supplémentaire de maturation pour lui permettre d'acquérir l'extension G-riche caractéristique (Wellinger et al. 1996). De plus, il a été montré que la création de ces extensions a également lieu dans une souche de levure ayant perdu la télomérase (Dionne et Wellinger 1996), suggérant que la formation des extensions lors de la phase S ne dépend pas uniquement de l'enzyme télomérase (Dionne et Wellinger 1996). Avec l'utilisation de plasmides possédant ou non une origine de réplication, il a été montré que la création des extensions G-riches est dépendante du passage d'une fourche de réplication (Dionne et Wellinger 1998). Ces données suggèrent donc un lien entre la machinerie de réplication conventionnelle de l'ADN et le maintien des télomères (Dionne et Wellinger 1998).

Étant donné l'importance et l'ubiquité des extensions G-riches, il a toujours été supposé que ces extensions devaient être présentes aux télomères de la levure en dehors de la phase S, malgré le manque de données à ce sujet. Une protéine impliquée depuis

longtemps dans le maintien de l'intégrité des télomères chez la levure est Cdc13p (Garvik et al. 1995; Lydall et Weinert 1995; Nugent et al. 1996). Il a été montré que cette protéine lie l'ADN télomérique simple-brin G-riche de levure *in vitro* (Nugent et al. 1996; Lin et Zakian 1996; Hughes et al. 2000; Lin et al. 2001; Mitton-Fry et al. 2002) et aussi qu'elle peut lier les télomères de levure *in vivo* (Bourns et al. 1998; Diede et Gottschling 2001; Tsukamoto et al. 2001; Taggart et al. 2002). Par la suite, les rôles de Cdc13p dans les deux fonctions essentielles associées aux télomères ont été mieux définis. Premièrement, son rôle dans la protection des télomères a été établi suite aux observations de sa liaison spécifique aux extensions simple-brin et à la formation d'un complexe protéique avec Stn1p et Ten1p (Grandin et al. 1997; Wang et al. 2000; Pennock et al. 2001; Grandin et al. 2001). Deuxièmement, son rôle dans la réplication des télomères a été établi grâce à sa liaison avec la protéine Est1p, (Bianchi et al. 2004; Evans et Lundblad 1999; Grandin et al. 2000; Pennock et al. 2001; Qi et Zakian 2000; Chandra et al. 2001; Taggart et al. 2002; Schramke et al. 2004), et aussi grâce à sa liaison à la sous-unité catalytique de l'ADN polymérase α (Pol1p) (Qi et Zakian 2000). Plusieurs modèles ont été suggérés récemment pour expliquer les diverses interactions protéiques de Cdc13p et son rôle complexe dans la régulation des télomères (Takata et al. 2005; Hsu et al. 2004; Jia et al. 2004; Grandin et Charbonneau 2003; Chandra et al. 2001; Bianchi et al. 2004; Meier et al. 2001; Pennock et al. 2001), revues dans (Smogorzewska et de Lange 2004; Kanoh et Ishikawa 2003; Lustig 2001). Tous ces modèles reposent sur la supposition que des extensions simple-brin sont présentes aux télomères de la levure à l'extérieur de la phase S.

D'autres études font état de résultats concernant la liaison de protéines aux télomères de levure, selon les phases du cycle cellulaire (Taggart et al. 2002; Smith et al.

2003; Schramke et al. 2004; Fisher et al. 2004). En utilisant des essais de ChIP (chromatin immunoprecipitation), le groupe de Zakian a suivi la localisation cellulaire des protéines Est1p, Est2p et Cdc13p *in vivo* (Taggart et al. 2002). Leurs résultats montrent que la protéine Cdc13p est associée aux télomères durant les phases G1 et S (Taggart et al. 2002). Ils montrent également la localisation de la protéine Est2p aux télomères en phases G1 et S (Taggart et al. 2002), même si la télomérase n'est pas active en phase G1 (Diede et Gottschling 1999; Marcand et al. 2000). Un autre groupe a obtenu des résultats similaires concernant la localisation de la protéine Est2p en utilisant également des essais de ChIP (Smith et al. 2003). Les résultats d'une seconde étude du même groupe utilisant des essais de ChIP, ont montré que la présence du complexe hétérodimérique Ku est essentielle pour l'association de Est2p au télomère en phase G1 (Fisher et al. 2004). La distribution de la protéine Est1p aux télomères est différente et elle se retrouve fortement associée aux télomères en phase S tardive, au moment de l'action de la télomérase *in vivo*. Les résultats de liaison de protéines obtenus par ChIP ne permettent pas de différencier si la liaison des protéines a lieu à l'extrémité simple-brin ou à une région double-brin du télomère. Cependant, étant donné la caractérisation exhaustive de la liaison de la protéine Cdc13p à l'ADN simple-brin télomérique G-riche de levure *in vitro* et *in vivo* (Nugent et al. 1996; Lin et Zakian 1996; Hughes et al. 2000; Lin et al. 2001; Anderson et al. 2002; Mitton-Fry et al. 2002; Mitton-Fry et al. 2004; Bourns et al. 1998; Diede et Gottschling 2001; Tsukamoto et al. 2001; Taggart et al. 2002), il est facile de supposer que les résultats obtenus concernent la liaison de Cdc13p au simple-brin télomérique selon les phases du cycle cellulaire (Taggart et al. 2002).

Cependant, aucune évidence à ce moment n'était disponible concernant la présence de cette extension ailleurs qu'en phase S (Wellinger et al. 1993b; Wellinger et al. 1993a).

Des études ont également permis d'obtenir des informations sur la séquence 3' terminale retrouvée sur un télomère de levure. Suite à l'amplification par PCR et au séquençage d'un télomère spécifique génétiquement marqué, les résultats montrent une dégénérescence et une grande hétérogénéité des télomères de levure, au niveau de leur séquence et de leur longueur (Forstemann et al. 2000). Il a été montré que chez la levure, au contraire de ce qui a été observé chez les ciliés (Fan et Price 1997; Jacob et al. 2001) et chez les humains (Sfeir et al. 2005), l'extrémité simple-brin est irrégulière. Aussi, au contraire des observations faites chez les ciliés et l'humain, la toute dernière base en 3' du télomère de levure n'est pas conservée, retrouvant aléatoirement un G ou un T à cette position (Forstemann et al. 2000), et aucun patron régulier de séquence ne peut être déduit des 20 dernières bases séquencées. Ces observations concernant l'irrégularité des séquences télomériques peuvent être reliées au fait que la télomérase de levure est peu processive et peut facilement déraiser lors de l'ajout de séquences terminales (Cohn et Blackburn 1995; Petrov et al. 1998; Peng et al. 2001; Forstemann et Lingner 2001; Forstemann et al. 2003; Seto et al. 2003; Bosoy et Lue 2004; revue dans Lue 2004). Ces observations sur l'hétérogénéité de la séquence télomérique de levure et le manque de conservation des 20 dernières bases du télomère peuvent questionner la possibilité qu'une protéine spécifique telle Cdc13p puisse se lier directement à des extensions télomériques non conservées (voir plus loin), et aussi questionner la formation possible de t-loops aux chromosomes de levure. Pour qu'il y ait formation de t-loops aux télomères, il faut présumer la possibilité d'un appariement adéquat entre un important nombre de bases du brin

G-riche et du brin C-riche complémentaire, ce qui semble impossible chez la levure à cause de la dégénérescence des séquences télomériques. De plus, la présence de ces t-loops est jusqu'à maintenant observée dans des organismes où de longues extensions simple-brin sont présentes aux télomères (Griffith et al. 1999; Murti et Prescott 1999; Munoz-Jordan et al. 2001; Nikitina et Woodcock 2004), ce qui n'est pas le cas chez la levure.

La présence d'extensions G-riche aux télomères de levure en phase G1 du cycle cellulaire a été démontrée et par le fait même les données expérimentales obtenues auparavant par différents groupes ont été rendues possibles et cohérentes (Larrivée et al. 2004). Des subtilités apportées à la méthode d'hybridation dans le gel ont permis la détection des extensions simple-brin des cellules de type sauvage qui auparavant était impossible à cause du faible signal généré par les courtes extensions télomériques présentes aux télomères endogènes des cellules asynchrones. Les extensions simple-brin télomériques dans des cellules de type sauvage détectées par cette méthode sont plus courtes que 16 bases (Larrivée et al. 2004). Étant donné que les extensions télomériques mesurées ici contiennent des séquences variables (Forstemann et al. 2000), elles ne sont pas totalement complémentaires à la sonde télomérique utilisée. Pour remédier à ce problème, nous avons comparé toutes les séquences théoriques possibles pour des extensions télomériques de 16 nts et la séquence de la sonde télomérique utilisée. Toutes les séquences possibles qui peuvent être présentes aux télomères de levure possèdent au moins 13 nucléotides qui sont parfaitement complémentaires à la sonde C₁₋₃A utilisée (C. LeBel et R.J. Wellinger, données non-publiées). Ainsi, les données suggèrent que les

extensions télomériques sur un plasmide linéaire contenu dans les levures de type sauvage en phase G1 sont de 13 à 16 nucléotides (Larrivée et al. 2004).

Une approche d'extension d'amorce modifiée a également été élaborée pour mesurer la taille des extensions télomériques présentes *in vivo* aux extrémités des chromosomes de levure. Étant donné l'hétérogénéité et l'irrégularité des séquences télomériques de levure, les méthodes de mesure utilisées chez d'autres organismes ayant des répétitions régulières sont impossibles à utiliser. La procédure a été appliquée à des télomères purifiés de levures de type sauvage arrêtées en phase G1, et la taille des extensions télomériques récupérées est d'environ 12 à 14 nucléotides (Larrivée et al. 2004). L'intensité des bandes correspondant aux produits d'extension a été quantifiée, et il ressort qu'au moins la moitié des extensions simple-brin des levures en G1 seraient entre 9 à 12 nucléotides.

Nous avons vérifié si cette extension simple-brin détectée aux télomères de levures en phase G1 est une caractéristique partagée par tous les télomères. Il en ressort que la vaste majorité, et possiblement tous les télomères de levure possèdent une extension 3' simple-brin en G1 (figure 4) (Larrivée et al. 2004). La présence de cette extension simple-brin chez toutes les espèces étudiées laissait présager que celle-ci devait être également présente à tous les télomères de la levure, ce qui a été démontré hors de tout doute avec nos résultats (Larrivée et al. 2004).

Il a été montré que la séquence minimale pour permettre la liaison de la protéine Cdc13p *in vitro* est 11 nucléotides, 5'-dGTGTGGGTGTG-3' (Hughes et al. 2000), et que la protéine peut aussi lier un substrat de 10 nucléotides TG₁₋₃ (Lin et Zakian 1996). Il a été montré que des variants naturels de la séquence minimale de liaison sont liés avec une

affinité comparable (Anderson et al. 2003). De plus, certains acides aminés de la région du DBD de Cdc13p (Hughes et al. 2000; Lin et al. 2001; Anderson et al. 2002; Mitton-Fry et al. 2002), reconnaissent les nucléotides 5'-GTGT qui sont requis pour la haute affinité de liaison (Anderson et al. 2003). La taille des extensions télomériques présentes aux télomères de levures de type sauvage en phase G1 permet donc la liaison de la protéine Cdc13p aux extrémités simple-brin, et par la suite, le recrutement des autres protéines impliquées dans la protection des extrémités et dans la réplication. La liaison de Cdc13p à l'acide nucléique est accomplie grâce à un domaine de type OB (OB fold) présent dans le DBD de Cdc13p et à une grande boucle d'environ 30 acides aminés qui sont responsables des contacts avec l'ADN (Mitton-Fry et al. 2002; Anderson et al. 2003; Mitton-Fry et al. 2004). Les domaines OB fold sont habituellement de petits motifs structuraux utilisés pour la liaison avec de courts ligands (2-5 nts) (Murzin 1993). La présence de la boucle d'acides aminés aidant à la liaison du substrat de 11 nucléotides télomériques permet d'expliquer les divergences entre les domaines OB de Cdc13p et les domaines OB usuels (Mitton-Fry et al. 2002). La protéine TEBP qui lie l'extension simple-brin chez *O. nova* peut lier un substrat simple-brin de 12 nts (Horvath et al. 1998; Horvath et Schultz 2001). Ce type de domaine OB fold a été identifié chez plusieurs protéines permettant la protection des extensions simple-brin chez différentes espèces (Mitton-Fry et al. 2002; Anderson et al. 2003; Theobald et al. 2003; Theobald et Schultz 2003; Theobald et Wuttke 2004; Mitton-Fry et al. 2004). Le domaine OB fold détecté dans Cdc13p chez la levure *S. cerevisiae* est similaire en structure à ceux détectés chez les autres espèces, mais possède une séquence primaire totalement différente (Mitton-Fry et al. 2002; Anderson et al. 2003; Theobald et al. 2003; Theobald et Schultz 2003;

Theobald et Wuttke 2004; Mitton-Fry et al. 2004). Cependant, l'usage commun d'un domaine OB fold pour l'interaction entre les protéines télomériques et l'ADN terminal montre la conservation et l'importance des mécanismes de protection des extrémités chez les eucaryotes.

En ce qui concerne l'irrégularité des séquences terminales de levure et la liaison de la protéine Cdc13p, il a été montré que Cdc13p lie spécifiquement et avec une grande affinité l'ADN simple-brin G-riche présent chez la levure, mais qu'elle est aussi capable de lier les séquences télomériques simple-brin G-riche de *O. nova* et de l'humain, cependant de façon moins efficace (Nugent et al. 1996; Lin et Zakian 1996). Son mode de liaison est différent de celui observé pour l'hétérodimère TEBP de *O. nova* qui requiert une extension télomérique 3', qui se retrouve ensevelie dans le complexe (Gottschling et Zakian 1986; Horvath et al. 1998). Il est aussi différent du mode de liaison de la protéine Pot1p de *S. pombe* dans lequel l'ADN est impliqué dans les interactions intramoléculaires cruciales pour la spécificité (Lei et al. 2002; Lei et al. 2003). En effet, il a été observé que l'ADN lié par Cdc13p serait enroulé autour de la surface du domaine OB fold contenu dans le DBD de Cdc13p (Anderson et al. 2003; Mitton-Fry et al. 2004). Cette conformation de l'ADN serait semblable à celles observées lors de la liaison de la protéine SSB chez les procaryotes (Raghunathan et al. 2000), et lors de la liaison de RPA chez les eucaryotes (Bochkarev et al. 1997), qui lient l'ADN sans la présence de séquence spécifique. Fonctionnellement, la protéine Cdc13p doit être capable de reconnaître l'ADN télomérique simple-brin de séquence dégénérée qui est présent aux télomères de levure. Ceci peut être accompli par la liaison de la séquence 5'-GTGT (Anderson et al. 2003; Mitton-Fry et al. 2004) qui agirait comme un motif

signature présent dans l'extension télomérique simple-brin de tous les télomères de levure (Forstemann et al. 2000; Forstemann et Lingner 2001). Cette tolérance de séquences variables d'ADN simple-brin par Cdc13p peut être expliquée par un mécanisme semblable à celui qui est aussi observé chez *O. nova*, appelé « nucleotide shuffling » (Theobald et Schultz 2003). En effet, le complexe TEBP responsable de la liaison des extensions simple-brin chez *O. nova* permet la liaison de substrat non-apparenté à la séquence habituellement reconnue par le complexe (Theobald et Schultz 2003). La conformation globale du complexe reste à peu près identique à celle observée lors de la liaison aux substrats habituels, mais l'ADN simple-brin subit des réarrangements structuraux, pouvant mener à « l'expulsion » de certains nucléotides hors du complexe formé, ce qui a été appelé le « nucleotide shuffling » (Theobald et Schultz 2003). Ce mode de liaison pourrait aussi être présent chez la levure et permettrait la liaison des séquences télomériques dégénérées présentes aux extrémités 3' terminales. Les modes de liaison alternatifs observés pour le DBD de Cdc13p avec des oligonucléotides comportant des substitutions permettent de supporter cette interprétation (Mitton-Fry et al. 2004).

Le complexe MRX serait l'exonucléase 5'-3' responsable de la génération et du maintien des extensions 3' simple-brin chez la levure

Lors de la réplication de l'ADN par la machinerie de réplication conventionnelle, l'enlèvement de l'amorce ARN du dernier fragment d'Okazaki créera une courte extension simple-brin à une des extrémités du chromosome (brin retardé) (Lingner et al. 1995; Greider 1996; Zakian 1996). Par contre, de l'autre côté du chromosome, la

synthèse du brin avancé laissera un bout franc (Lingner et al. 1995; Greider 1996; Zakian 1996). Chez la levure *S. cerevisiae*, l'analyse des intermédiaires de réplication formés au niveau des télomères montre l'apparition d'extensions G-riches d'au moins 30 bases lors de la phase S tardive, qui correspond au moment de la réplication des télomères (Wellinger et al. 1993a; Wellinger et al. 1993b; Marcand et al. 2000). Ces extensions sont retrouvées à chacune des extrémités d'un plasmide linéaire (Wellinger et al. 1993b), et sont aussi générées à la fin de la phase S dans des cellules dépourvues de télomérase (Dionne et Wellinger 1996; Wellinger et al. 1996). La création de ces extensions n'est donc pas due à l'ajout de séquences télomériques par la télomérase et ces évidences ont mené à la suggestion d'une nouvelle étape dans la réplication des télomères, soit la présence d'une exonucléase 5'-3' qui permettrait la formation des extensions simple-brin requises aux deux extrémités des chromosomes (Wellinger et al. 1996). Très récemment, la présence d'une nucléase spécifique qui pourrait agir sur le brin C-riche dans les cellules humaines a été proposée (Sfeir et al. 2005). Cette activité serait nécessaire pour la formation du substrat de la télomérase (Lingner et al. 1995; Lingner et Cech 1996) et pour la génération de la structure terminale requise pour la liaison des protéines permettant la protection des chromosomes. Il a été observé que la génération de l'extension simple-brin G-riche est dépendante du passage de la machinerie de réplication (Dionne et Wellinger 1998). Ainsi, cette exonucléase pourrait être associée à la machinerie de réplication et en son absence, les deux extrémités du chromosome pourraient être traitées différemment suite à la réplication : l'extension simple-brin faite suite à la réplication du brin retardé serait de taille normale, alors que l'extrémité faite

suite à la synthèse du brin avancé donnerait une extrémité franche qui n'est pas traitée de façon normale.

Plusieurs gènes encodent des exonucléases 5'-3' chez la levure qui peuvent être des candidats pour cette fonction. Des analyses de comparaison de séquences ont permis d'identifier un motif conservé chez plusieurs enzymes avec une activité exonucléolytique (Harrington et Lieber 1994). Ce motif a servi pour identifier des gènes encodant des exonucléases potentielles dans le génome de *S. cerevisiae* (I. Dionne, S. Gravel, J. Parenteau et R.J. Wellinger, données non-publiées). Suite à ces considérations de séquences, 5 candidats pouvant être responsables de l'activité exonucléase générant les extensions simple-brin G-riches chez la levure ont été récupérés (I. Dionne, S. Gravel, J. Parenteau et R.J. Wellinger, données non-publiées). Ces candidats potentiels sont EXOI (Fiorentini et al. 1997; Tishkoff et al. 1997), RAD17, RAD24 (Lydall et Weinert 1995), TRM2/NUD1/NUC2 (Chow et Resnick 1987; Moore et al. 1993), DIN7 (Mieczkowski et al. 1997; Fikus et al. 2000) et YEN1 (Johnson et al. 1998). Suite à la délétion des gènes encodant ces exonucléases potentielles dans des souches de levure, aucun effet n'a été observé sur la génération des extensions simple-brin ou la taille des télomères (I. Dionne, S. Gravel, J. Parenteau et R.J. Wellinger, données non-publiées), suggérant que ces protéines ne contiennent pas l'activité nucléolytique générant les extensions simple-brin.

Le complexe MRE11/RAD50/XRS2 (MRX) chez la levure ou le complexe MRE11/RAD50/NBS1 (MRN) chez les cellules humaines, semble être un complexe conservé à travers l'évolution, qui possède plusieurs fonctions essentielles chez différents organismes. Il est impliqué dans la réparation des cassures d'ADN double-brin, et chez la levure, dans le processus de NHEJ et de recombinaison homologue (revue dans

D'Amours et Jackson 2002). Chez les vertébrés, le complexe est important pour préserver l'intégrité des fourches durant la réplication de l'ADN (Costanzo et al. 2001). Le complexe est impliqué dans le « checkpoint » de la phase S et également au niveau des télomères chez la levure, les mammifères et les plantes (revue dans Haber 1998; D'Amours and Jackson 2002). L'absence de l'une des trois protéines du complexe mène à un phénotype de télomères très courts dans la levure (Kironmai et Muniyappa 1997; Boulton et Jackson 1998). De plus, il a été montré que la présence du complexe est importante pour le maintien de la longueur des télomères en absence de la télomérase, ce qui suggère qu'il pourrait être impliqué dans la maturation des télomères pour leur réplication par la télomérase (Nugent et al. 1998; Ritchie et Petes 2000). Dans les cellules humaines, le complexe MRN interagit avec la protéine TRF2 et la présence du complexe aux télomères peut être détectée en phase S (Zhu et al. 2000; Lombard et Guarente 2000). La protéine Mre11p est une endonucléase spécifique pour le simple-brin et une exonucléase 3'-5' (Furuse et al. 1998; Usui et al. 1998). Cependant, une mutation dans le domaine nucléase de Mre11p n'a pas d'effet sur la taille des télomères, indiquant que le domaine nucléase de Mre11p ne serait pas impliqué dans le maintien des télomères (Moreau et al. 1999; Tsukamoto et al. 2001). La protéine Rad50p peut se lier à l'ADN de façon dépendante de l'ATP (Raymond et Kleckner 1993). L'homologue humain de Xrs2p, NBS1, confère une activité hélicase au complexe MRN humain (Paull et Gellert 1999).

Le complexe MRX possède une activité exonucléase 5'-3' *in vivo*, qui est nécessaire pour son activité dans la recombinaison des cassures d'ADN double-brin. Une donnée conflictuelle survient à ce sujet : le groupe de Gottschling a montré que le

complexe MRX est nécessaire à la formation des extensions G-riches médiées par la télomérase et donc à la liaison de la protéine Cdc13p à ces extensions (Diede et Gottschling 1999), mais le groupe de Zakian a montré que le complexe MRX n'est pas essentiel pour la formation des extensions G-riches et la liaison de la protéine Cdc13p, et qu'il serait plutôt requis pour le recrutement de l'activité télomérase aux télomères (Tsukamoto et al. 2001). La différence dans les données générées par les deux groupes peut être expliquée par la disparité des méthodes utilisées. Dans la première étude, l'addition *de novo* de séquences télomériques par la télomérase est suivie sur un télomère artificiel. Celui-ci est créé suite à une coupure par l'endonucléase HO dont la production est induite par un promoteur répondant au galactose (Diede et Gottschling 1999). Les cellules observées sont en phase M, et il est possible que la coupure dans la séquence d'ADN qui est induite par l'endonucléase HO soit vue comme une cassure d'ADN double-brin, qui doit être à tout prix réparée par la cellule. Comme le complexe MRX est essentiel pour la réparation de ces cassures, il est possible que le télomère artificiel soit « réparé » plutôt que traité comme un vrai télomère. Ceci expliquerait que la protéine Cdc13p ne se lie pas efficacement à ce télomère artificiel. Ces expériences auraient été plus valides si un télomère artificiel avait été créé de façon autre que suite à une coupure d'ADN double-brin, par exemple en utilisant la méthode de recombinaison avec l'enzyme Flp (Marcand et al. 1999; Marcand et al. 2000), qui n'affecte pas l'intégrité du télomère. La deuxième étude analyse les télomères naturels de cellules non-synchronisées, par ChIP (Tsukamoto et al. 2001). Il aurait été plus valide de faire les études dans des cellules synchronisées et relâchées dans le cycle cellulaire pour vérifier si la quantité de protéine Cdc13p liée aux télomères est comparable dans les cellules de type

sauvage et dans les cellules *mre11Δ*. Ainsi, en comparant ces données, l'implication du complexe MRX dans la création des extensions simple-brin demeure discutable.

Les résultats que nous avons obtenus montrent l'implication du complexe MRX au niveau des télomères. L'ADN génomique isolé d'une souche de levure *mre11Δ* en phase G1 du cycle cellulaire possède des extensions télomériques plus courtes que celles des levures de type sauvage (figure 1, Larrivée et al. 2004). La quantification des signaux des extensions simple-brin permet de montrer qu'il y a environ 1,3 à 2 fois moins d'ADN simple-brin chez les cellules *mre11Δ* que chez les cellules de type sauvage (figure 1, Larrivée et al. 2004). Comme les extensions détectées aux télomères des cellules de type sauvage par hybridation dans le gel sont de 12-14 bases, ceci implique que celles présentes dans les cellules *mre11Δ* soient de 6-10 bases. Des analyses de la longueur des extensions simple-brin ont montré que les extensions présentes dans les cellules *mre11Δ* en phase G1 sont d'environ 8 bases (figure 3, Larrivée et al. 2004). Cependant, des produits de 9 à 12 bases de longueur sont aussi détectables, mais les produits d'extension entre 2 et 8 bases de longueur sont plus abondants dans les cellules *mre11Δ* que les produits les plus longs. Ainsi, les souches analysées peuvent avoir des extensions simple-brin de 12-14 bases, mais la fraction des télomères avec une extension simple-brin de cette longueur est grandement réduite dans les cellules *mre11Δ* (Larrivée et al. 2004). Ces données semblent confirmer celles obtenues par Diede et Gottschling (Diede et Gottschling 2001) et montrent l'implication du complexe MRX dans la génération des extensions télomériques simple-brin. Il est possible que des extensions de longueur normale soient présentes à l'extrémité du chromosome qui a été formée par la réplication du brin retardé (figure 33). La réplication du brin avancé peut donner une extrémité de

chromosome qui est franche et qui requiert le complexe MRX pour générer une extension simple-brin G-riche de longueur appropriée pour la liaison de la protéine Cdc13p. Dans les cellules de type sauvage, le complexe MRX se chargerait de cette tâche. Cependant, la création des extensions simple-brin est tellement importante pour les télomères que la cellule ne peut s'en passer et des activités exonucléases redondantes au complexe MRX sont présentes afin d'assurer la formation de ces extensions. Ceci permet d'expliquer la présence d'extensions télomériques de même longueur que celles observées dans les cellules de type sauvage (12-14 bases), dans les cellules *mre11Δ*. Ces extensions seraient dues à l'enlèvement de l'amorce ARN lors de la réplication du brin retardé. Les extensions les plus courtes observées dans les cellules *mre11Δ* seraient créées lors de la réplication du brin avancé, par une activité redondante au complexe MRX. Si aucune exonucléase redondante au complexe MRX n'était présente dans les cellules *mre11Δ*, il est possible que les extensions simple-brin ne soient pas établies et/ou présentes à une extrémité du chromosome. Plusieurs protéines pourraient faire partie des exonucléases redondantes au complexe MRX, et un candidat potentiel pourrait être la protéine Dna2p. En effet, il a été récemment montré que l'homologue de cette protéine chez la levure *S. pombe* est requis pour la génération des extensions télomériques simple-brin (Tomita et al. 2004).

L'ensemble de ces données suggère un traitement différent de chacune des extrémités d'un chromosome lors de la réplication. Des observations en ce sens ont été faites chez les cellules humaines et les plantes, où la présence d'extensions télomériques asymétriques selon l'extrémité du chromosome a été observée (Wright et al. 1997; Riha et al. 2000). De plus, il a également été montré que la réplication des télomères par la

synthèse du brin avancé et retardé mène à des structures terminales différentes en absence de la protéine TRF2 (Bailey et al. 2001). L'absence de TRF2 dans les cellules de mammifères mène à des fusions entre télomères, mais seulement ceux issus de la synthèse du brin avancé (Bailey et al. 2001), ce qui suggère qu'un défaut dans la protection des extrémités de chromosomes affecterait seulement ces télomères. Les télomères issus de la synthèse du brin retardé possèderaient une extension simple-brin et une structure adéquate.

Les analyses ont aussi permis de montrer que la dynamique de formation des extensions simple-brin n'est pas affectée dans les cellules *mre11Δ* (figure 5, Larrivée et al. 2004) et que c'est seulement la taille des extensions simple-brin qui varie dans les cellules *mre11Δ* (Larrivée et al. 2004). Ces données sont en accord avec celles qui montrent que, dans les cellules *mre11Δ*, la formation des extensions simple-brin n'est pas complètement abolie mais la structure télomérique est altérée (Diede et Gottschling 2001). La liaison de la protéine Cdc13p ne semble pas affectée par la présence des courtes extensions télomériques (Tsukamoto et al. 2001), et il est possible que la longueur minimale pour la liaison de la protéine Cdc13p *in vivo* soit moins que les 11 bases requises pour la liaison maximale *in vitro* (Hughes et al. 2000). Il est aussi possible que la protéine Cdc13p puisse lier les courtes extensions télomériques de façon alternative *in vivo*. Il a été montré récemment que le complexe MRX est recruté aux télomères de levures tard en phase S, et que ce recrutement est requis pour le positionnement adéquat de la kinase Mec1p aux télomères (Takata et al. 2005). À son tour, la protéine Mec1p contribuerait au recrutement des protéines Cdc13p et Est1p aux extrémités des chromosomes (Takata et al. 2005). Ces données suggèrent un assemblage

hiérarchique des protéines télomériques en phase S, qui est amorcé par l'association du complexe MRX aux télomères (Takata et al. 2005).

Nos données supportent l'implication du complexe MRX dans la préparation adéquate de l'ADN télomérique, qui doit être établie correctement pour le restant du cycle cellulaire et aussi pour la reconnaissance et l'action par la télomérase (Nugent et al. 1998; Larrivée et al. 2004; Takata et al. 2005). Ce rôle du complexe MRX au niveau des télomères pourrait être conservé entre les espèces. Il a été montré que l'équivalent protéique de Xrs2p chez l'humain (NBS1) s'associe de façon transitoire à la fin de la phase S, avec un complexe télomérique composé des protéines RAD50/MRE11/TRF2 (Zhu et al. 2000). Cette association spécifique a été suggérée comme étant nécessaire pour la préparation des télomères à la réplication ou pour un événement post-réplicatif, comme l'établissement des extensions télomériques appropriées et/ou la formation des t-loops (Zhu et al. 2000).

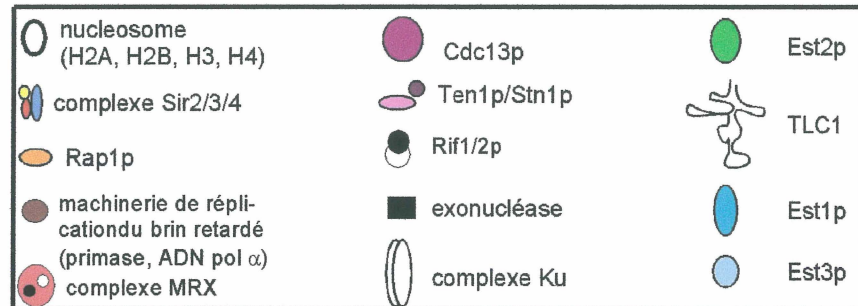
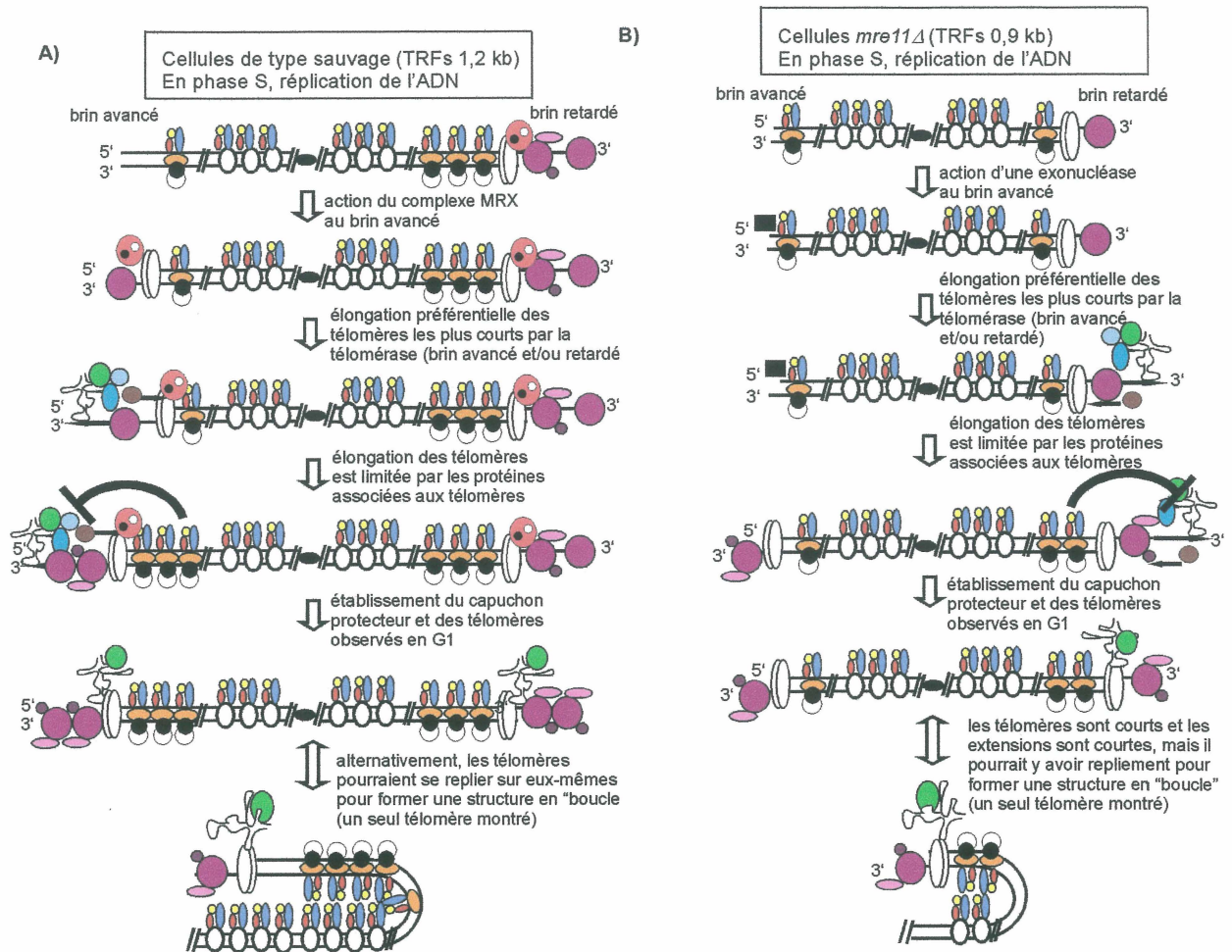


Figure 33. Modèle pour la réplication des télomères de levure et pour l'établissement de la structure télomérique observée en phase G1. A) Dans les cellules de type sauvage, les télomères sont d'environ 1,2 kb. La réplication de l'ADN devrait donner un bout franc (brin avancé) et un bout avec une extension simple-brin (brin retardé). Le complexe MRX est l'exonucléase responsable de la création de l'extension simple-brin au brin avancé. La télomérase agit ensuite préférentiellement sur le télomère le plus court afin d'ajouter des séquences télomériques, et la machinerie de réplication du brin retardé suit pour rallonger le brin complémentaire. La télomérase peut agir sur les deux brins (avancé et/ou retardé), sur un seul des deux brins ou encore ne pas agir du tout, selon les cycles. Dans la figure, l'allongement du brin avancé est illustré. Suite à l'élongation, il y a recrutement des protéines associées aux télomères, ce qui permet de stopper l'élongation et de reformer le capuchon protecteur aux extrémités des chromosomes. Le complexe MRX peut être impliqué dans l'étape d'élongation ou de formation du capuchon. Chez la levure, la structure de t-loop n'a pas été observée, mais une structure semblable impliquant un repliement de la chromatine télomérique est possible. Les structures des télomères observées en phase G1 ont été créées lors de la phase S précédente. B) Dans les cellules *mre11Δ*, les télomères sont très courts (0,9 kb). La réplication de l'ADN devrait donner un bout franc (brin avancé) et un bout avec une extension simple-brin (brin retardé). Le complexe MRX étant absent, une exonucléase redondante serait responsable de la création de l'extension simple-brin au brin avancé. Étant donné que tous les télomères sont courts, la télomérase devrait agir sur le télomère le plus court afin d'ajouter des séquences télomériques et d'assurer la viabilité cellulaire. La télomérase peut agir sur les deux brins (avancé et/ou retardé), sur un seul des deux brins ou encore ne pas agir, selon les cycles. Les extensions rallongées par la télomérase peuvent être d'une longueur similaire aux extensions présentes dans les cellules de type sauvage, mais les autres extensions restent très courtes, permettant la liaison de Cdc13p et la protection des extrémités. La légende des protéines est la même pour les figures A et B.

Analyse des extensions télomériques simple-brin dans des lignées de cellules

humaines avec une méthode de T-OLA modifiée

Le modèle mis en place par Hayflick et collaborateurs, montrant que le noyau de la cellule peut gérer le nombre de divisions cellulaires à performer avant l'entrée en sénescence, implique l'existence d'un « réplicomètre » (Muggleton-Harris et Hayflick 1976; Wright et Hayflick 1975; Hayflick 1975). Les études ont montré que le raccourcissement des télomères pouvait être un des mécanismes moléculaires responsables de la sénescence répllicative, et que ce phénomène serait donc un « réplicomètre ». Selon ce modèle, la longueur des télomères, qui diminue avec chaque ronde de division cellulaire à cause du processus de réplication comme tel, dicte le nombre de divisions possibles que la cellule peut accomplir en absence de moyens pour maintenir la taille de ses télomères. Lorsque les télomères atteignent une taille critique pour la cellule, un signal serait envoyé pour mettre fin à la prolifération. Cependant, de nombreuses observations indiquent que le raccourcissement des télomères n'est pas le seul mécanisme impliqué dans la sénescence répllicative. Selon les différents modèles

expérimentaux, il y a d'autres mécanismes qui jouent un rôle dans la limite de la capacité répliquative des cellules. L'accumulation graduelle de dommages causés par le stress engendré par la culture des cellules et le stress environnemental pourraient induire un arrêt de prolifération qui serait le résultat d'une accumulation de dommages (Stewart et Weinberg 2002; Wright et Shay 2002). Des études dans cette voie proposent que les télomères ne seraient que des « détecteurs » de l'instabilité génétique produite par d'autres mécanismes, dont le stress oxydatif (revue dans Von Zglinicki 2003). Un autre modèle a également été mis en place, et implique l'existence des télomères sous deux formes possibles, soit un état de protection et un état déprotégé (Blackburn 2000). Ce ne serait donc pas le raccourcissement des télomères comme tel qui déclencherait la sénescence mais bien la perte de la protection adéquate aux télomères (Blackburn 2001). Ainsi, l'hypothèse la plus plausible présentement est qu'en absence de télomérase active dans les cellules, il y a un raccourcissement progressif des télomères dû aux différentes rondes de répllication et à l'accumulation de stress chronique, autant dans la portion double-brin que dans l'extension simple-brin des télomères. Ce raccourcissement peut causer l'effondrement de la structure télomérique normale et envoyer certains signaux pour que la cellule entre en sénescence. De cette hypothèse naît le besoin de pouvoir mesurer ou analyser les extensions simple-brin au niveau moléculaire et avec une résolution augmentée.

La dynamique des télomères dans les cellules humaines est une composante critique qui a plusieurs implications dans des processus cellulaires importants, tels le vieillissement et la carcinogénèse. Au fil des ans, différentes méthodes ont été développées dans le but de mieux étudier et de caractériser les télomères de cellules

humaines. Ces méthodes incluent l'hybridation dans le gel adaptée pour les télomères de cellules humaines (développée auparavant chez la levure (Dionne et Wellinger 1996)), différentes méthodes de visualisation par hybridation *in situ* de sondes télomériques spécifiques (FISH, Q-FISH, CO-FISH) (revue dans Baerlocher et al. 2002; Baerlocher et Lansdorp 2004; Kapoor et Telford 2004), une méthode appelée STELA (single telomere length analysis) basée sur l'amplification par PCR d'un télomère unique (Baird et al. 2003), une méthode appelée PENT (primer-extension/nick translation) pour la détection et la quantification des extensions simple-brin se basant sur l'hybridation d'amorces spécifiques et l'extension par la Taq polymérase (Makarov et al. 1997), une méthode d'extension d'amorce (McElligott et Wellinger 1997), une méthode de visualisation des télomères par microscopie électronique (Wright et al. 1997; Griffith et al. 1999), une méthode de mesure de l'extension télomérique basée sur l'hybridation et la ligation d'oligos complémentaires à cette extension (T-OLA) (Cimino-Reale et al. 2001) et une méthode basée sur la protection des extensions télomériques par liaison de protéines spécifiques et visualisation des extensions simple-brin (Chai et al. 2005).

Les mesures du simple-brin télomérique des cellules humaines ont donné des résultats différents selon les méthodes utilisées. Les résultats obtenus par extension d'amorce montrent que des extensions simple-brin de plus de 45 bases sont présentes aux extrémités des chromosomes de cellules normales et transformées, donnée qui est confirmée par les résultats d'hybridation dans le gel (McElligott et Wellinger 1997). Les résultats obtenus par PENT et analyse par gel en deux dimensions, montrent que plus de 80% des télomères issus de cellules normales ou transformées possèdent une extension simple-brin qui serait de 130 à 210 bases (Makarov et al. 1997). En utilisant des protéines

liant spécifiquement le simple-brin et la microscopie électronique pour examiner les extensions télomériques, il a été montré que les extensions sont en moyenne de 200 ± 75 nucléotides chez les fibroblastes (Wright et al. 1997). Par la suite, la méthode de T-OLA a montré que la longueur des extensions G-riches peut être de 24 à plus de 400 nucléotides, avec une majorité de celles-ci (56-62%) faisant moins de 90 nucléotides et les extensions de 108-270 nucléotides comptant pour seulement 37% de toutes les extensions, et ce pour tous les types cellulaires examinés (Cimino-Reale et al. 2001).

Nous avons apporté des modifications à la méthode de T-OLA (figure 27) (Cimino-Reale et al. 2001) dans le but d'augmenter la précision des résultats obtenus pour la mesure des extensions télomériques simple-brin. Cette méthode permet de déterminer la taille relative des extensions télomériques présentes sur une population de télomères, mais ne permet pas la détermination de la taille absolue des extensions, ni la taille précise d'une extension télomérique particulière. Les résultats obtenus en utilisant cette méthode de T-OLA modifiée sont en accord avec ceux obtenus pour la mesure des extensions télomériques faites avec les autres méthodes. La taille des extensions simple-brin des télomères de cellules transformées (HeLa, ECR 293, HCT 116, HA 5) peut mesurer jusqu'à 240 nucléotides, avec une majorité des extensions située entre 60 et 130 nucléotides. Avec cette méthode, les bandes des courts produits semblent majoritaires dans le patron de bandes et leur forte intensité suggère la présence de plus de produits de ligation de cette taille. Cependant, le dosage de leur intensité par rapport à l'intensité totale de toutes les bandes du patron aurait permis de quantifier la contribution de chacune des différentes bandes.

Nous avons examiné le patron des extensions simple-brin dans les cellules ALT. Ce type cellulaire semblait intéressant à étudier car aucune donnée n'est disponible sur les extensions télomériques présentes à leurs télomères. Le mécanisme impliqué pour le maintien de la longueur des télomères dans les cellules ALT n'est pas complètement compris. Plusieurs évidences suggèrent que l'addition de séquences télomériques a lieu par un mécanisme de recombinaison plutôt que par l'action de la télomérase (Henson et al. 2002; Bryan et al. 1997a; Bryan et al. 1997b). À cet effet, il a été montré qu'une séquence particulière présente sur un télomère dans une lignée cellulaire ALT, peut être diffusée et retrouvée sur des télomères additionnels (Dunham et al. 2000). Étant donné que des événements de recombinaison de l'ADN impliquent la présence d'extension simple-brin, il nous semblait intéressant de mieux caractériser l'extension simple-brin dans les cellules ALT. Ce mécanisme de recombinaison pourrait également impliquer la présence de cercles extrachromosomaux contenant des séquences télomériques (appelés t-circles ou cercles-t), comme ce fut observé chez la levure, la grenouille, et dans certaines lignées de cellules humaines (Tomaska et al. 2004a). Chez la levure *K. lactis*, la présence de ces cercles-t peut promouvoir le maintien des télomères par recombinaison (Tomaska et al. 2000; Natarajan et McEachern 2002; Natarajan et al. 2003; Underwood et al. 2004). Chez *S. cerevisiae*, la présence de cercles contenant des séquences télomériques et des éléments Y' a été observée (Horowitz et Haber 1985; Louis et Haber 1990), pouvant éventuellement aider au maintien des télomères.

Tout comme l'activation de la télomérase dans les cellules (Kim et al. 1994; Bodnar et al. 1998), cette façon alternative de maintenir la taille des télomères dans les cellules ALT peut mener au développement de certains cancers (Reddel et Bryan 2003).

Cependant, ces formes de cancers seraient intraitables par des inhibiteurs de télomérase. Les cellules ALT possèdent des télomères qui sont de longueur très hétérogène, variant de très courts (<1 kb) à très longs (>20 kb) (revue dans Tomaska et al. 2004a) (C. Le Bel et R.J. Wellinger, données non-publiées). Cette hétérogénéité est générée par la combinaison de mécanismes menant au raccourcissement des télomères comme normalement observé dans des cellules qui sont négatives pour l'expression de la télomérase, et des événements de rallongement et de raccourcissement rapides (Murnane et al. 1994; Perrem et al. 2001). Une autre caractéristique des cellules ALT est la présence de corps PMLs (promyelocytic leukemia) contenant de l'ADN télomérique, des protéines qui se lient aux télomères (TRF1 et TRF2) et des protéines ayant une fonction dans les mécanismes de recombinaison (Yeager et al. 1999). Les corps PMLs sont des structures dynamiques de réponse au stress, retrouvées dans la plupart des cellules et ayant plusieurs fonctions (revue dans Dellaire et Bazett-Jones 2004). Cependant, des corps PMLs ayant le contenu spécifique de ceux retrouvés dans les cellules ALT existent seulement dans ce type de cellules et sont inexistantes dans des cellules mortelles ou des lignées cellulaires qui sont positives pour l'expression de la télomérase. Ces corps PMLs ont été nommés APBs (ALT-associated PML bodies) (Yeager et al. 1999) et sont retrouvés dans ~ 5% des cellules d'une population asynchrone de cellules ALT en division. Les APBs pourraient servir d'entrepôts pour les molécules impliquées dans le mécanisme ALT (Yeager et al. 1999). En ce sens, il a été montré par imagerie qu'un télomère peut effectivement être en contact avec un APB puis en sortir (Molenaar et al. 2003), suggérant ainsi que le contact entre eux puisse permettre le mécanisme ALT.

Les résultats préliminaires obtenus pour la mesure des extensions simple-brin de cellules GM 847 (ALT) montrent que les extensions seraient plus courtes que celles observées dans des cellules transformées ou des cellules normales avec une taille maximale d'environ 100-120 nucléotides. Cette taille serait la limite pour permettre la formation des t-loops aux extrémités des chromosomes (Griffith et al. 1999; Stansel et al. 2001). Même si les télomères des cellules ALT peuvent être très longs et hétérogènes, l'extension simple-brin présente aux télomères de ces cellules semble passablement plus courte que celle observée dans les autres types cellulaires. Il est possible que ce soit le mécanisme de maintien des télomères par recombinaison comme tel qui génère des extensions télomériques plus courtes que celles générées par l'addition de séquences par la télomérase. Il est également possible que ce mécanisme de maintien des télomères ne requière pas la présence d'une longue extension télomérique. La présence de t-loops dans les cellules ALT demeure indéterminée, et un autre mode de protection des extrémités pourrait être important dans ces cellules.

Maintien des télomères dans des cellules ALT : modèles possibles

Le maintien des télomères dans les cellules ALT pourrait avoir lieu grâce à certains mécanismes précis impliquant la recombinaison. La recombinaison entre des séquences télomériques dans des cellules ALT mène à un rallongement de la longueur des télomères mais aussi à une instabilité télomérique (Reddel 2003), qui peut inclure des événements de délétion de séquences télomériques similaires aux événements de TRD (telomere rapid deletion) qui sont observés chez la levure (Lustig 2003). Également dans les cellules ALT, il peut y avoir présence de fragments extra-chromosomaux contenant

des répétitions télomériques (Ogino et al. 1998; Tokutake et al. 1998). Des fragments extrachromosomaux semblables ont été retrouvés dans des cellules de patients souffrant d'ataxia-telangiesctasia et chez des souris déficientes en protéine ATM ($ATM^{-/-}$) (Hande et al. 2001). Il est possible que la présence de cercles composés de répétitions télomériques agisse dans les cellules comme un « réservoir » de séquences, permettant ainsi la synthèse de longs télomères par les ADN polymérases conventionnelles (Lindstrom et al. 2002).

In vitro, il a été montré que des cercles aussi petits que 36 à 60 nucléotides de répétitions télomériques (6 à 10 répétitions télomériques humaines) peuvent servir de substrat pour la synthèse de télomères humains par les ADN polymérases (Hartig et Kool 2004). De plus, deux récents ouvrages montrent que les cellules ALT possèdent des cercles-t, suggérant un rôle potentiel de ces cercles dans la réplication des télomères en absence de la télomérase (Cesare et Griffith 2004; Wang et al. 2004). Cependant, aucune donnée *in vivo* dans les cellules humaines ne montre que les cercles-t peuvent être utilisés comme substrat pour l'amplification des répétitions télomériques. Pour que ce mécanisme soit efficace *in vivo*, il doit y avoir présence d'une extension télomérique simple-brin qui s'hybride de façon complémentaire à un nombre déterminé de bases afin d'enclencher la réplication sous forme de cercle roulant (revue dans Khan 2005; Tomaska et al. 2004a; Novick 1998). Une autre façon de former ces cercles-t serait que l'extension télomérique simple-brin présente aux télomères des cellules ALT envahisse la région double-brin du télomère, formant une structure de t-loop. Ce t-loop pourrait être vu comme un intermédiaire de recombinaison et la résolution de cette structure pourrait mener à un raccourcissement des télomères et/ou à la formation d'un cercle-t (figure 34).

Les conditions exigées pour que ces mécanismes d'allongement des télomères aient lieu dans la cellule impliquent que les télomères doivent être dans un état « déprotégé » pour permettre l'accès à l'extension simple-brin. Comme il a été montré que le télomère le plus court est critique pour la viabilité de la cellule et la stabilité des chromosomes (Hemann et al. 2001), il est possible que ce télomère subisse un rallongement préférentiel, afin de maintenir la viabilité cellulaire. Il est également possible que de courts télomères avec des extensions simple-brin courtes soient incapables de former les t-loops, et que par conséquent, ils soient engagés dans des mécanismes alternatifs de maintien des télomères. Ceci est aussi possible dans des cellules qui expriment la télomérase, car il a été montré qu'il peut y avoir co-existence des 2 modes de maintien des télomères (télomérase et ALT) dans différents types de cellules (Perrem et al. 2001; Reddel 2003; Tomaska et al. 2004a). Par contre, étant donné qu'il est plus facile de détecter l'activité de la télomérase que les mécanismes alternatifs moins bien connus et compris, cette voie de maintien a probablement été sous-estimée pour sa contribution dans le maintien des télomères. Chez la levure, la télomérase a été localisée aux télomères par ChIP durant le cycle cellulaire (Taggart et al. 2002) et il a été suggéré qu'elle soit une composante du capuchon télomérique (Blackburn 2000). Chez les cellules humaines, il est possible que la télomérase soit impliquée dans la protection des extrémités et la formation des t-loops, ce qui rendrait le télomère inaccessible à d'autres mécanismes de dégradation ou d'élongation. Les mécanismes ALT pourraient donc s'établir plus facilement dans des cellules où la télomérase n'est pas présente aux télomères et où les télomères sont accessibles et doivent être allongés. La présence d'extensions plus courtes dans les cellules ALT suggère que ces mécanismes peuvent être

moins efficaces que l'addition de séquences télomériques par la télomérase pour maintenir la taille de l'extension simple-brin, mais que ces mécanismes seraient efficaces pour le maintien de la longueur globale des télomères. Il est possible que ces mécanismes de recombinaison utilisés pour le maintien des télomères exigent des extensions télomériques moins longues que celles requises pour former une t-loop, et que par conséquent, ces structures ne soient pas présentes dans les cellules ALT. Une fois le mode de recombinaison enclenché par un télomère très court pour permettre de maintenir la viabilité de la cellule, ce mode de maintien deviendrait plus facile et serait continué dans la cellule.

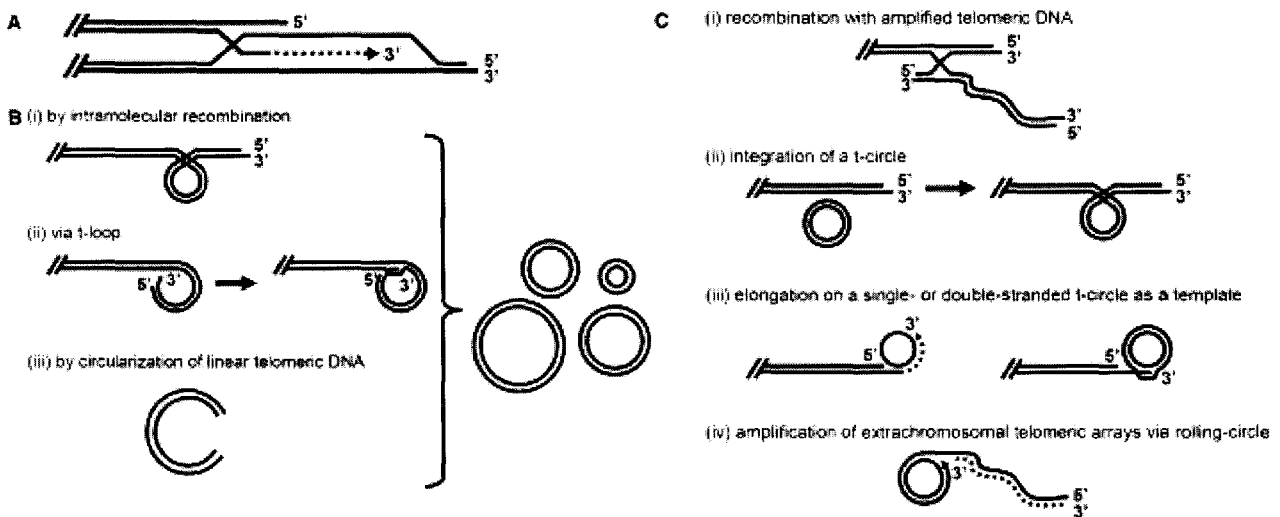


Figure 34. Mécanismes possibles pour le maintien des télomères dans les cellules ALT. A) Élongation des télomères par un mécanisme de recombinaison, qui implique l'invasion par l'extension télomérique simple-brin de la séquence double-brin d'un autre télomère, et l'utilisation du brin complémentaire comme substrat pour l'élongation. B) Mécanismes possibles de formation des cercles-t. (i) Recombinaison homologe intramoléculaire entre des répétitions télomériques (ii) Recombinaison homologe intramoléculaire initiée à une t-loop (iii) Ligation de fragments télomériques linéaires. C) Mécanismes possibles d'élongation des télomères médiés par de l'ADN télomérique extrachromosomal. (i) Recombinaison avec un fragment télomérique linéaire. (ii) Recombinaison avec un cercle-t. (iii) Allongement d'un télomère par synthèse selon le mécanisme de cercle-roulant utilisant un cercle-t comme substrat pour l'élongation. (iv) Amplification de séquences télomériques extrachromosomales sur des cercles-t, par le mécanisme de réplication de cercle-roulant. La recombinaison de séquences extrachromosomales avec des télomères pourrait mener à l'élongation des télomères. Tiré de Tomaska et al. (2004a).

CONCLUSIONS

Mes travaux ont permis de montrer les interactions cellulaires variées de la protéine hnRNP A1, et de caractériser ses différents domaines impliqués dans la modulation de l'épissage alternatif. Mes travaux ont montré les rôles similaires joués par les protéines hnRNP A1 et hnRNP A2 dans la sélection des sites d'épissage, sur un transcrit possédant des sites de liaison optimaux.

Mes travaux de caractérisation des extensions télomériques ont permis de montrer la présence d'extensions simple-brin courtes durant la phase G1 du cycle cellulaire chez la levure, de même que l'implication du complexe MRX dans la création et le maintien de ces extensions simple-brin. Des résultats préliminaires sur la longueur des extensions télomériques simple-brin présentes dans des cellules ALT suggèrent certaines différences entre les structures et les modes de maintien des télomères chez les cellules humaines.

REMERCIEMENTS

Je peux maintenant prendre le temps de dire **MERCI** à tous ceux et celles qui m'ont aidée, de près ou de loin, en le sachant ou non, avec leurs paroles, leurs gestes ou encore leurs silences.

Tous (tes) et chacun(e), je voudrais que vous sachiez que votre support continu, vos encouragements permanents, et encore beaucoup plus que ce que vous pouvez imaginer, m'ont beaucoup aidée.

Merci à Dr. Raymund Wellinger et Dr. Benoit Chabot, de m'avoir acceptée dans leur laboratoire respectif et de m'avoir fait confiance avec certains projets, d'avoir toujours démontré une rigueur scientifique et un esprit critique, mais ouvert, et surtout de m'avoir enseigné le plaisir de « faire de la bonne science » tout en gardant le côté humain.

Merci à Carole Picard et aux membres passés et présents des laboratoires du département.

Merci à Dre. Lea Harrington, Dr. Martin Bisaillon, et Dre. Claudine Rancourt d'avoir accepté de constituer le jury de cette thèse.

Je me dois de remercier ceux qui ont été présents pour moi durant mes longues études;

Merci à ma famille, Maman, Papa et JP; vous avez toujours su me dire à quel point vous étiez fiers de moi. Votre immense support à tous les niveaux, vos encouragements sincères et soutenus ont toujours été une source de motivation pour moi. Malgré la distance, je sais que vous pensez à moi...Merci de m'avoir aidée à aller jusqu'au bout!

Merci à Jo, mon « biggest fan », pour sa motivation contagieuse dont j'avais besoin, son écoute et ses conseils aux moments cruciaux, son immense cœur, et surtout, merci d'avoir été avec moi à travers tout ça.

Merci à Terry et Huguette, de m'avoir constamment appuyée et encouragée, et merci pour toute votre générosité.

Merci à Lise, de m'avoir enseigné la signification profonde de « SFA » et d'avoir été là, tout simplement, comme une mère adoptive; à Michel, pour les questions et l'intérêt manifesté face à mes projets; à grand-maman Tétreau, pour la fierté exprimée, les appuis constants et les encouragements assidus.

Merci à Raymonde, pour son authenticité et sa persévérance exemplaires; à Vézeau, qui est parti trop tôt, pour son courage face aux épreuves; à grand-maman LeBel, pour sa soif de connaissances et ses félicitations constantes.

Merci à Nicole R. d'avoir suivi mon cheminement et de m'avoir toujours exprimé sa fierté.

Merci à mes amis (es) du labo, du RECMUS, du département, de randonnée, de ski, d'escalade, de vélo de montagne, de gym, de cinéma, de magasinage, etc., et mes amis (es) tout court.

Merci au CRSNG/NSERC pour son appui financier.

RÉFÉRENCES

- Abdul-Manan, N., S.M. O'Malley, et K.R. Williams. (1996). Origins of binding specificity of the A1 heterogeneous nuclear ribonucleoprotein. *Biochemistry* **35**, 3545-54.
- Abdul-Manan, N. et K.R. Williams. (1996). hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base-specific binding. *Nucleic Acids Res* **24**, 4063-70.
- Adams, M.D., R.S. Tarnag, et D.C. Rio. (1997). The alternative splicing factor PSI regulates P-element third intron splicing in vivo. *Genes Dev* **11**, 129-38.
- Allsopp, R.C., H. Vaziri, C. Patterson, S. Goldstein, E.V. Younghi, A.B. Futcher, et al. (1992). Telomere length predicts replicative capacity of human fibroblasts. *Proc Natl Acad Sci U S A* **89**, 10114-8.
- Anderson, E.M., W.A. Halsey, et D.S. Wuttke. (2002). Delineation of the high-affinity single-stranded telomeric DNA-binding domain of *Saccharomyces cerevisiae* Cdc13. *Nucleic Acids Res* **30**, 4305-13.

- . (2003). Site-directed mutagenesis reveals the thermodynamic requirements for single-stranded DNA recognition by the telomere-binding protein Cdc13. *Biochemistry* **42**, 3751-8.
- Armbruster, B.N., C.M. Linardic, T. Veldman, N.P. Bansal, D.L. Downie, et C.M. Counter. (2004). Rescue of an hTERT mutant defective in telomere elongation by fusion with hPot1. *Mol Cell Biol* **24**, 3552-61.
- Baerlocher, G.M. et P.M. Lansdorp. (2004). Telomere length measurements using fluorescence in situ hybridization and flow cytometry. *Methods Cell Biol* **75**, 719-50.
- Baerlocher, G.M., J. Mak, T. Tien, et P.M. Lansdorp. (2002). Telomere length measurement by fluorescence in situ hybridization and flow cytometry: tips and pitfalls. *Cytometry* **47**, 89-99.
- Bai, Y., D. Lee, T. Yu, et L.A. Chasin. (1999). Control of 3' splice site choice in vivo by ASF/SF2 and hnRNP A1. *Nucleic Acids Res* **27**, 1126-34.
- Bailey, S.M., M.N. Cornforth, A. Kurimasa, D.J. Chen, et E.H. Goodwin. (2001). Strand-specific postreplicative processing of mammalian telomeres. *Science* **293**, 2462-5.

- Baird, D.M., J. Rowson, D. Wynford-Thomas, et D. Kipling. (2003). Extensive allelic variation and ultrashort telomeres in senescent human cells. *Nat Genet* **33**, 203-7.
- Baker, B.S. (1989). Sex in flies: the splice of life. *Nature* **340**, 521-4.
- Bartel, P., C.T. Chien, R. Sternglanz, et S. Fields. (1993). Elimination of false positives that arise in using the two-hybrid system. *Biotechniques* **14**, 920-4.
- Baudin, A., O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, et C. Cullin. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **21**, 3329-30.
- Baumann, P. et T.R. Cech. (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171-5.
- Baumann, P., E. Podell, et T.R. Cech. (2002). Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing. *Mol Cell Biol* **22**, 8079-87.
- Bell, L.R., J.I. Horabin, P. Schedl, et T.W. Cline. (1991). Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in *Drosophila*. *Cell* **65**, 229-39.

- Bell, L.R., E.M. Maine, P. Schedl, et T.W. Cline. (1988). Sex-lethal, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55**, 1037-46.
- Ben-David, Y., M.R. Bani, B. Chabot, A. De Koven, et A. Bernstein. (1992). Retroviral insertions downstream of the heterogeneous nuclear ribonucleoprotein A1 gene in erythroleukemia cells: evidence that A1 is not essential for cell growth. *Mol Cell Biol* **12**, 4449-55.
- Berget, S.M. (1995). Exon recognition in vertebrate splicing. *J Biol Chem* **270**, 2411-4.
- Bériault, V., J.F. Clement, K. Levesque, C. Lebel, X. Yong, B. Chabot, et al. (2004). A late role for the association of hnRNP A2 with the HIV-1 hnRNP A2 response elements in genomic RNA, Gag, and Vpr localization. *J Biol Chem* **279**, 44141-53.
- Berman, J., C.Y. Tachibana, et B.K. Tye. (1986). Identification of a telomere-binding activity from yeast. *Proc Natl Acad Sci U S A* **83**, 3713-7.

Beyer, A.L., M.E. Christensen, B.W. Walker, et W.M. LeStourgeon. (1977).

Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell* **11**, 127-38.

Biamonti, G., M.T. Bassi, L. Cartegni, F. Mechta, M. Buvoli, F. Cobianchi, et al. (1993).

Human hnRNP protein A1 gene expression. Structural and functional characterization of the promoter. *J Mol Biol* **230**, 77-89.

Bianchi, A., S. Negrini, et D. Shore. (2004). Delivery of yeast telomerase to a DNA break

depends on the recruitment functions of Cdc13 and Est1. *Mol Cell* **16**, 139-46.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev*

Biochem **72**, 291-336.

Blackburn, E.H. (2000). Telomere states and cell fates. *Nature* **408**, 53-6.

-. (2001). Switching and signaling at the telomere. *Cell* **106**, 661-73.

Blackburn, E.H. et P.B. Challoner. (1984). Identification of a telomeric DNA sequence in

Trypanosoma brucei. *Cell* **36**, 447-57.

Blackburn, E.H. et J.G. Gall. (1978). A tandemly repeated sequence at the termini of the

extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J Mol Biol* **120**, 33-53.

- Blanchette, M. et B. Chabot. (1997). A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. *RNA* **3**, 405-19.
- . (1999). Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J* **18**, 1939-52.
- Blencowe, B.J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* **25**, 106-10.
- Bochkarev, A., R.A. Pfuetzner, A.M. Edwards, et L. Frappier. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* **385**, 176-81.
- Bodnar, A.G., M. Ouellette, M. Frolkis, S.E. Holt, C.P. Chiu, G.B. Morin, et al. (1998). Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349-52.
- Bosoy, D. et N.F. Lue. (2004). Yeast telomerase is capable of limited repeat addition processivity. *Nucleic Acids Res* **32**, 93-101.
- Boue, S., I. Letunic, et P. Bork. (2003). Alternative splicing and evolution. *Bioessays* **25**, 1031-4.

- Boulton, S.J. et S.P. Jackson. (1996). *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J* **15**, 5093-103.
- . (1998). Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J* **17**, 1819-28.
- Bourns, B.D., M.K. Alexander, A.M. Smith, et V.A. Zakian. (1998). Sir proteins, Rif proteins, and Cdc13p bind *Saccharomyces* telomeres in vivo. *Mol Cell Biol* **18**, 5600-8.
- Brett, D., H. Pospisil, J. Valcarcel, J. Reich, et P. Bork. (2002). Alternative splicing and genome complexity. *Nat Genet* **30**, 29-30.
- Brow, D.A. (2002). Allosteric cascade of spliceosome activation. *Annu Rev Genet* **36**, 333-60.
- Bryan, T.M., A. Englezou, L. Dalla-Pozza, M.A. Dunham, et R.R. Reddel. (1997a). Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat Med* **3**, 1271-4.

- Bryan, T.M., A. Englezou, J. Gupta, S. Bacchetti, et R.R. Reddel. (1995). Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J* **14**, 4240-8.
- Bryan, T.M., L. Marusic, S. Bacchetti, M. Namba, et R.R. Reddel. (1997b). The telomere lengthening mechanism in telomerase-negative immortal human cells does not involve the telomerase RNA subunit. *Hum Mol Genet* **6**, 921-6.
- Bryan, T.M. et R.R. Reddel. (1997). Telomere dynamics and telomerase activity in in vitro immortalised human cells. *Eur J Cancer* **33**, 767-73.
- Burd, C.G. et G. Dreyfuss. (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J* **13**, 1197-204.
- Buvoli, M., G. Biamonti, P. Tsoulfas, M.T. Bassi, A. Ghetti, S. Riva, et al. (1988). cDNA cloning of human hnRNP protein A1 reveals the existence of multiple mRNA isoforms. *Nucleic Acids Res* **16**, 3751-70.

- Buvoli, M., F. Cobianchi, M. Bestagno, M.T. Bassi, G. Biamonti, et S. Riva. (1990a). A second A1-type protein is encoded by the human hnRNP A1 gene. *Mol Biol Rep* **14**, 83-4.
- Buvoli, M., F. Cobianchi, G. Biamonti, et S. Riva. (1990b). Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. *Nucleic Acids Res* **18**, 6595-600.
- Buvoli, M., F. Cobianchi, et S. Riva. (1992). Interaction of hnRNP A1 with snRNPs and pre-mRNAs: evidence for a possible role of A1 RNA annealing activity in the first steps of spliceosome assembly. *Nucleic Acids Res* **20**, 5017-25.
- Caceres, J.F., S. Stamm, D.M. Helfman, et A.R. Krainer. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* **265**, 1706-9.
- Caputi, M., A. Mayeda, A.R. Krainer, et A.M. Zahler. (1999). hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO J* **18**, 4060-7.

Cartegni, L., S.L. Chew, et A.R. Krainer. (2002). Listening to silence and understanding

nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-98.

Cartegni, L., M. Maconi, E. Morandi, F. Cobianchi, S. Riva, et G. Biamonti. (1996).

hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J Mol Biol* **259**, 337-48.

Casas-Finet, J.R., J.D. Smith, Jr., A. Kumar, J.G. Kim, S.H. Wilson, et R.L. Karpel.

(1993). Mammalian heterogeneous ribonucleoprotein A1 and its constituent domains. Nucleic acid interaction, structural stability and self-association. *J Mol Biol* **229**, 873-89.

Cech, T.R. (2004). Beginning to understand the end of the chromosome. *Cell* **116**, 273-9.

Celotto, A.M. et B.R. Graveley. (2001). Alternative splicing of the Drosophila Dscam

pre-mRNA is both temporally and spatially regulated. *Genetics* **159**, 599-608.

Cesare, A.J. et J.D. Griffith. (2004). Telomeric DNA in ALT cells is characterized by

free telomeric circles and heterogeneous t-loops. *Mol Cell Biol* **24**, 9948-57.

Chabot, B. (1996). Directing alternative splicing: cast and scenarios. *Trends Genet* **12**,

472-8.

- Chabot, B., M. Blanchette, I. Lapierre, et H. La Branche. (1997). An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol Cell Biol* **17**, 1776-86.
- Chabot, B., C. LeBel, S. Hutchison, F.H. Nasim, et M.J. Simard. (2003). Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA. *Prog Mol Subcell Biol* **31**, 59-88.
- Chai, W., J.W. Shay, et W.E. Wright. (2005). Human telomeres maintain their overhang length at senescence. *Mol Cell Biol* **25**, 2158-68.
- Chakhparonian, M. et R.J. Wellinger. (2003). Telomere maintenance and DNA replication: how closely are these two connected? *Trends Genet* **19**, 439-46.
- Chan, C.S. et B.K. Tye. (1983). Organization of DNA sequences and replication origins at yeast telomeres. *Cell* **33**, 563-73.
- Chandler, S.D., A. Mayeda, J.M. Yeakley, A.R. Krainer, et X.D. Fu. (1997). RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proc Natl Acad Sci U S A* **94**, 3596-601.

- Chandra, A., T.R. Hughes, C.I. Nugent, et V. Lundblad. (2001). Cdc13 both positively and negatively regulates telomere replication. *Genes Dev* **15**, 404-14.
- Chien, C.T., P.L. Bartel, R. Sternglanz, et S. Fields. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* **88**, 9578-82.
- Chow, T.Y. et M.A. Resnick. (1987). Purification and characterization of an endonuclease from *Saccharomyces cerevisiae* that is influenced by the RAD52 gene. *J Biol Chem* **262**, 17659-67.
- Cimino-Reale, G., E. Pascale, E. Alvino, G. Starace, et E. D'Ambrosio. (2003). Long telomeric C-rich 5'-tails in human replicating cells. *J Biol Chem* **278**, 2136-40.
- Cimino-Reale, G., E. Pascale, E. Battiloro, G. Starace, R. Verna, et E. D'Ambrosio. (2001). The length of telomeric G-rich strand 3'-overhang measured by oligonucleotide ligation assay. *Nucleic Acids Res* **29**, E35.
- Clouet d'Orval, B., Y. d'Aubenton Carafa, P. Sirand-Pugnet, M. Gallego, E. Brody, et J. Marie. (1991a). RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. *Science* **252**, 1823-8.

Clouet d'Orval, B., Y. d'Aubenton-Carafa, J.M. Brody, et E. Brody. (1991b).

Determination of an RNA structure involved in splicing inhibition of a muscle-specific exon. *J Mol Biol* **221**, 837-56.

Cobianchi, F., R.L. Karpel, K.R. Williams, V. Notario, et S.H. Wilson. (1988).

Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. Large-scale overproduction in *Escherichia coli* and cooperative binding to single-stranded nucleic acids. *J Biol Chem* **263**, 1063-71.

Cohn, M. et E.H. Blackburn. (1995). Telomerase in yeast. *Science* **269**, 396-400.

Colgin, L.M., K. Baran, P. Baumann, T.R. Cech, et R.R. Reddel. (2003). Human POT1 facilitates telomere elongation by telomerase. *Curr Biol* **13**, 942-6.

Collins, K. (1996). Structure and function of telomerase. *Curr Opin Cell Biol* **8**, 374-80.

-. (1999). Ciliate telomerase biochemistry. *Annu Rev Biochem* **68**, 187-218.

Collins, K. et J.R. Mitchell. (2002). Telomerase in the human organism. *Oncogene* **21**, 564-79.

Coschigano, K.T. et P.C. Wensink. (1993). Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila*. *Genes Dev* **7**, 42-54.

- Costanzo, V., K. Robertson, M. Bibikova, E. Kim, D. Grieco, M. Gottesman, et al. (2001). Mre11 protein complex prevents double-strand break accumulation during chromosomal DNA replication. *Mol Cell* **8**, 137-47.
- Côté, J. et B. Chabot. (1997). Natural base-pairing interactions between 5' splice site and branch site sequences affect mammalian 5' splice site selection. *RNA* **3**, 1248-61.
- Counter, C.M., A.A. Avilion, C.E. LeFeuvre, N.G. Stewart, C.W. Greider, C.B. Harley, et al. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J* **11**, 1921-9.
- Counter, C.M., M. Meyerson, E.N. Eaton, et R.A. Weinberg. (1997). The catalytic subunit of yeast telomerase. *Proc Natl Acad Sci U S A* **94**, 9202-7.
- Coyne, R.S., D.L. Chalker, et M.C. Yao. (1996). Genome downsizing during ciliate development: nuclear division of labor through chromosome restructuring. *Annu Rev Genet* **30**, 557-78.
- Cristofalo, V.J., R.G. Allen, R.J. Pignolo, B.G. Martin, et J.C. Beck. (1998). Relationship between donor age and the replicative lifespan of human cells in culture: a reevaluation. *Proc Natl Acad Sci U S A* **95**, 10614-9.

- Cristofalo, V.J., A. Lorenzini, R.G. Allen, C. Torres, et M. Tresini. (2004). Replicative senescence: a critical review. *Mech Ageing Dev* **125**, 827-48.
- Dallaire, F., S. Dupuis, S. Fiset, et B. Chabot. (2000). Heterogeneous nuclear ribonucleoprotein A1 and UP1 protect mammalian telomeric repeats and modulate telomere replication in vitro. *J Biol Chem* **275**, 14509-16.
- D'Amours, D. et S.P. Jackson. (2002). The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. *Nat Rev Mol Cell Biol* **3**, 317-27.
- Daoud, R., M. Da Penha Berzaghi, F. Siedler, M. Hubener, et S. Stamm. (1999). Activity-dependent regulation of alternative splicing patterns in the rat brain. *Eur J Neurosci* **11**, 788-802.
- Das, R., Z. Zhou, et R. Reed. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell* **5**, 779-87.
- Datar, K.V., G. Dreyfuss, et M.S. Swanson. (1993). The human hnRNP M proteins: identification of a methionine/arginine-rich repeat motif in ribonucleoproteins. *Nucleic Acids Res* **21**, 439-46.

- Davis, C.A., L. Grate, M. Spingola, et M. Ares, Jr. (2000). Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res* **28**, 1700-6.
- de Bruin, D., S.M. Kantrow, R.A. Liberatore, et V.A. Zakian. (2000). Telomere folding is required for the stable maintenance of telomere position effects in yeast. *Mol Cell Biol* **20**, 7991-8000.
- de Bruin, D., Z. Zaman, R.A. Liberatore, et M. Ptashne. (2001). Telomere looping permits gene activation by a downstream UAS in yeast. *Nature* **409**, 109-113.
- de Lange, T., L. Shiue, R.M. Myers, D.R. Cox, S.L. Naylor, A.M. Killery, et al. (1990). Structure and variability of human chromosome ends. *Mol Cell Biol* **10**, 518-27.
- Dellaire, G., et D.P. Bazett-Jones. (2004). PML nuclear bodies: dynamic sensors of DNA damage and cellular stress. *Bioessays* **26**, 963-77.
- Dempsey, L.A., H. Sun, L.A. Hanakahi, et N. Maizels. (1999). G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. *J Biol Chem* **274**, 1066-71.

- Diede, S.J., et D.E. Gottschling. (1999). Telomerase-mediated telomere addition in vivo requires DNA primase and DNA polymerases alpha and delta. *Cell* **99**, 723-33.
- . (2001). Exonuclease activity is required for sequence addition and Cdc13p loading at a de novo telomere. *Curr Biol* **11**, 1336-40.
- Dignam, J.D., R.M. Lebovitz, et R.G. Roeder. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**, 1475-89.
- Ding, J., M.K. Hayashi, Y. Zhang, L. Manche, A.R. Krainer, et R.M. Xu. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* **13**, 1102-15.
- Dionne, I., et R.J. Wellinger. (1996). Cell cycle-regulated generation of single-stranded G-rich DNA in the absence of telomerase. *Proc Natl Acad Sci U S A* **93**, 13902-7.
- . (1998). Processing of telomeric DNA ends requires the passage of a replication fork. *Nucleic Acids Res* **26**, 5365-71.

Drabkin, H.A., J.D. West, M. Hotfilder, Y.M. Heng, P. Erickson, R. Calvo, et al. (1999).

DEF-3(g16/NY-LU-12), an RNA binding protein from the 3p21.3 homozygous deletion region in SCLC. *Oncogene* **18**, 2589-97.

Dreyfuss, G., M.J. Matunis, S. Pinol-Roma, et C.G. Burd. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **62**, 289-321.

Dunham, M.A., A.A. Neumann, C.L. Fasching, et R.R. Reddel. (2000). Telomere maintenance by recombination in human cells. *Nat Genet* **26**, 447-50.

Eperon, I.C., O.V. Makarova, A. Mayeda, S.H. Munroe, J.F. Caceres, D.G. Hayward, et al.(2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol* **20**, 8303-18.

Eperon, L.P., J.P. Estibeiro, et I.C. Eperon. (1986). The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. *Nature* **324**, 280-2.

Eperon, L.P., I.R. Graham, A.D. Griffiths, et I.C. Eperon. (1988). Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell* **54**, 393-401.

- Erlitzki, R., et M. Fry. (1997). Sequence-specific binding protein of single-stranded and unimolecular quadruplex telomeric DNA from rat hepatocytes. *J Biol Chem* **272**, 15881-90.
- Evans, S.K., et V. Lundblad. (1999). Est1 and Cdc13 as comediators of telomerase access. *Science* **286**, 117-20.
- . (2002). The Est1 subunit of *Saccharomyces cerevisiae* telomerase makes multiple contributions to telomere length maintenance. *Genetics* **162**, 1101-15.
- Eversole, A., et N. Maizels. (2000). In vitro properties of the conserved mammalian protein hnRNP D suggest a role in telomere maintenance. *Mol Cell Biol* **20**, 5425-32.
- Ewing, B., et P. Green. (2000). Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet* **25**, 232-4.
- Fan, X., et C.M. Price. (1997). Coordinate regulation of G- and C strand length during new telomere synthesis. *Mol Biol Cell* **8**, 2145-55.
- Faustino, N.A., et T.A. Cooper. (2003). Pre-mRNA splicing and human disease. *Genes Dev* **17**, 419-37.

Feng, J., W.D. Funk, S.S. Wang, S.L. Weinrich, A.A. Avilion, C.P. Chiu, et al. (1995).

The RNA component of human telomerase. *Science* **269**, 1236-41.

Fields, S., et O. Song. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-6.

Fikus, M.U., P.A. Mieczkowski, P. Koprowski, J. Rytka, E. Sledziowska-Gojska, et Z. Ciesla. (2000). The product of the DNA damage-inducible gene of *Saccharomyces cerevisiae*, DIN7, specifically functions in mitochondria. *Genetics* **154**, 73-81.

Fiorentini, P., K.N. Huang, D.X. Tishkoff, R.D. Kolodner, et L.S. Symington. (1997). Exonuclease I of *Saccharomyces cerevisiae* functions in mitotic recombination in vivo and in vitro. *Mol Cell Biol* **17**, 2764-73.

Fiset, S., et B. Chabot. (2001). hnRNP A1 may interact simultaneously with telomeric DNA and the human telomerase RNA in vitro. *Nucleic Acids Res* **29**, 2268-75.

Fisher, T.S., A.K. Taggart, et V.A. Zakian. (2004). Cell cycle-dependent regulation of yeast telomerase by Ku. *Nat Struct Mol Biol* **11**, 1198-205.

- Fitzgerald, M.S., T.D. McKnight, et D.E. Shippen. (1996). Characterization and developmental patterns of telomerase expression in plants. *Proc Natl Acad Sci U S A* **93**, 14422-7.
- Ford, L.P., J.W. Shay, et W.E. Wright. (2001). The La antigen associates with the human telomerase ribonucleoprotein and influences telomere length in vivo. *RNA* **7**, 1068-75.
- Ford, L.P., J.M. Suh, W.E. Wright, et J.W. Shay. (2000). Heterogeneous nuclear ribonucleoproteins C1 and C2 associate with the RNA component of human telomerase. *Mol Cell Biol* **20**, 9084-91.
- Ford, L.P., W.E. Wright, et J.W. Shay. (2002). A model for heterogeneous nuclear ribonucleoproteins in telomere and telomerase regulation. *Oncogene* **21**, 580-3.
- Forstemann, K., M. Hoss, et J. Lingner. (2000). Telomerase-dependent repeat divergence at the 3' ends of yeast telomeres. *Nucleic Acids Res* **28**, 2690-4.
- Forstemann, K. et J. Lingner. (2001). Molecular basis for telomere repeat divergence in budding yeast. *Mol Cell Biol* **21**, 7277-86.

Forstemann, K., A.J. Zaug, T.R. Cech, et J. Lingner. (2003). Yeast telomerase is specialized for C/A-rich RNA templates. *Nucleic Acids Res* **31**, 1646-55.

Frenck, R.W., Jr., E.H. Blackburn, et K.M. Shannon. (1998). The rate of telomere sequence loss in human leukocytes varies with age. *Proc Natl Acad Sci U S A* **95**, 5607-10.

Fu, X.D. et T. Maniatis. (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**, 437-41.

-. (1992a). The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proc Natl Acad Sci U S A* **89**, 1725-9.

-. (1992b). Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. *Science* **256**, 535-8.

Furuse, M., Y. Nagase, H. Tsubouchi, K. Murakami-Murofushi, T. Shibata, et K. Ohta. (1998). Distinct roles of two separable in vitro activities of yeast Mre11 in mitotic and meiotic recombination. *EMBO J* **17**, 6412-25.

- Garvik, B., M. Carson, et L. Hartwell. (1995). Single-stranded DNA arising at telomeres in *cdc13* mutants may constitute a specific signal for the RAD9 checkpoint. *Mol Cell Biol* **15**, 6128-38.
- Gattoni, R., D. Mahe, P. Mahl, N. Fischer, M.G. Mattei, J. Stevenin, et al. (1996). The human hnRNP-M proteins: structure and relation with early heat shock-induced splicing arrest and chromosome mapping. *Nucleic Acids Res* **24**, 2535-42.
- Gietz, D., A. St Jean, R.A. Woods, et R.H. Schiestl. (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res* **20**, 1425.
- Gietz, R.D., R.H. Schiestl, A.R. Willems, et R.A. Woods. (1995). Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* **11**, 355-60.
- Gorlach, M., C.G. Burd, et G. Dreyfuss. (1994a). The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J Biol Chem* **269**, 23074-8.
- . (1994b). The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp Cell Res* **211**, 400-7.

- Gottschling, D.E. et V.A. Zakian. (1986). Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell* **47**, 195-205.
- Grandin, N. et M. Charbonneau. (2003). The Rad51 pathway of telomerase-independent maintenance of telomeres can amplify TG1-3 sequences in *yku* and *cdc13* mutants of *Saccharomyces cerevisiae*. *Mol Cell Biol* **23**, 3721-34.
- Grandin, N., C. Damon, et M. Charbonneau. (2000). Cdc13 cooperates with the yeast Ku proteins and *stn1* To regulate telomerase recruitment. *Mol Cell Biol* **20**, 8397-408.
- . (2001). Ten1 functions in telomere end protection and length regulation in association with *Stn1* and *Cdc13*. *EMBO J* **20**, 1173-83.
- Grandin, N., S.I. Reed, et M. Charbonneau. (1997). *Stn1*, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with *Cdc13*. *Genes Dev* **11**, 512-27.
- Gravel, S. et R.J. Wellinger. (2002). Maintenance of double-stranded telomeric repeats as the critical determinant for cell viability in yeast cells lacking Ku. *Mol Cell Biol* **22**, 2182-93.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. *RNA* **6**, 1197-211.

-. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* **17**, 100-7.

Gray, J.T., D.W. Celandier, C.M. Price, et T.R. Cech. (1991). Cloning and expression of genes for the *Oxytricha* telomere-binding protein: specific subunit interactions in the telomeric complex. *Cell* **67**, 807-14.

Greider, C.W. (1996). Telomere length regulation. *Annu Rev Biochem* **65**, 337-65.

-. (1999). Telomeres do D-loop-T-loop. *Cell* **97**, 419-22.

Greider, C.W., et E.H. Blackburn. (1985). Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405-13.

-. (1987). The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**, 887-98.

Griffith, J.D., L. Comeau, S. Rosenfield, R.M. Stansel, A. Bianchi, H. Moss, et al. (1999). Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503-14.

- Grossel, M.J., H. Wang, B. Gadea, W. Yeung, et P.W. Hinds. (1999). A yeast two-hybrid system for discerning differential interactions using multiple baits. *Nat Biotechnol* **17**, 1232-3.
- Grunstein, M. (1997). Molecular model for telomeric heterochromatin in yeast. *Curr Opin Cell Biol* **9**, 383-7.
- Gure, A.O., N.K. Altorki, E. Stockert, M.J. Scanlan, L.J. Old, et Y.T. Chen. (1998). Human lung cancer antigens recognized by autologous antibodies: definition of a novel cDNA derived from the tumor suppressor gene locus on chromosome 3p21.3. *Cancer Res* **58**, 1034-41.
- Haber, J.E. (1998). The many interfaces of Mre11. *Cell* **95**, 583-6.
- Hamilton, B.J., C.M. Burns, R.C. Nichols, et W.F. Rigby. (1997). Modulation of AUUUA response element binding by heterogeneous nuclear ribonucleoprotein A1 in human T lymphocytes. The roles of cytoplasmic location, transcription, and phosphorylation. *J Biol Chem* **272**, 28732-41.

- Hamilton, B.J., E. Nagy, J.S. Malter, B.A. Arrick, et W.F. Rigby. (1993). Association of heterogeneous nuclear ribonucleoprotein A1 and C proteins with reiterated AUUUA sequences. *J Biol Chem* **268**, 8881-7.
- Hanamura, A., J.F. Caceres, A. Mayeda, B.R. Franza, Jr., et A.R. Krainer. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA* **4**, 430-44.
- Hande, M.P., A.S. Balajee, A. Tchirkov, A. Wynshaw-Boris, et P.M. Lansdorp. (2001). Extra-chromosomal telomeric DNA in cells from Atm(-/-) mice and patients with ataxia-telangiectasia. *Hum Mol Genet* **10**, 519-28.
- Harley, C.B. (1991). Telomere loss: mitotic clock or genetic time bomb? *Mutat Res* **256**, 271-82.
- Harley, C.B., A.B. Futcher, et C.W. Greider. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458-60.
- Harrington, J.J. et M.R. Lieber. (1994). Functional domains within FEN-1 and RAD2 define a family of structure-specific endonucleases: implications for nucleotide excision repair. *Genes Dev* **8**, 1344-55.

Harrington, L. (2003). Biochemical aspects of telomerase function. *Cancer Lett* **194**, 139-54.

Harrington, L., T. McPhail, V. Mar, W. Zhou, R. Oulton, M.B. Bass, et al. (1997). A mammalian telomerase-associated protein. *Science* **275**, 973-7.

Harrison, P.M., A. Kumar, N. Lang, M. Snyder, and M. Gerstein. (2002). A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res* **30**, 1083-90.

Hartig, J.S. et E.T. Kool. (2004). Small circular DNAs for synthesis of the human telomere repeat: varied sizes, structures and telomere-encoding activities. *Nucleic Acids Res* **32**, e152.

Hayflick, L. (1975). Current theories of biological aging. *Fed Proc* **34**, 9-13.

Hemann, M.T. et C.W. Greider. (1999). G-strand overhangs on telomeres in telomerase-deficient mouse cells. *Nucleic Acids Res* **27**, 3964-9.

Hemann, M.T., M.A. Strong, L.Y. Hao, et C.W. Greider. (2001). The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* **107**, 67-77.

Henderson, E. (1995). Telomere DNA structure. In *Telomeres* (ed. E.H. Blackburn and C.W. Greider), pp. 11-34. Cold Spring Harbor Laboratory Press, New York.

Henderson, E.R. et E.H. Blackburn. (1989). An overhanging 3' terminus is a conserved feature of telomeres. *Mol Cell Biol* **9**, 345-8.

Henson, J.D., A.A. Neumann, T.R. Yeager, et R.R. Reddel. (2002). Alternative lengthening of telomeres in mammalian cells. *Oncogene* **21**, 598-610.

Hertel, K.J., K.W. Lynch, E.C. Hsiao, E.H. Liu, et T. Maniatis. (1996). Structural and functional conservation of the Drosophila doublesex splicing enhancer repeat elements. *RNA* **2**, 969-81.

Hoffman, B.E. et P.J. Grabowski. (1992). U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev* **6**, 2554-68.

Horowitz, H. et J.E. Haber. (1985). Identification of autonomously replicating circular subtelomeric Y' elements in *Saccharomyces cerevisiae*. *Mol Cell Biol* **5**, 2369-80.

Horvath, M.P. et S.C. Schultz. (2001). DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex. *J Mol Biol* **310**, 367-77.

Horvath, M.P., V.L. Schweiker, J.M. Bevilacqua, J.A. Ruggles, et S.C. Schultz. (1998).

Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell* **95**, 963-74.

Hsu, C.L., Y.S. Chen, S.Y. Tsai, P.J. Tu, M.J. Wang, et J.J. Lin. (2004). Interaction of *Saccharomyces* Cdc13p with Pollp, Imp4p, Sir4p and Zds2p is involved in telomere replication, telomere maintenance and cell growth control. *Nucleic Acids Res* **32**, 511-21.

Huberman, J.A., L.D. Spotila, K.A. Nawotka, S.M. el-Assouli, et L.R. Davis. (1987). The in vivo replication origin of the yeast 2 microns plasmid. *Cell* **51**, 473-81.

Huffman, K.E., S.D. Levene, V.M. Tesmer, J.W. Shay, et W.E. Wright. (2000). Telomere shortening is proportional to the size of the G-rich telomeric 3'-overhang. *J Biol Chem* **275**, 19719-22.

Hughes, T.R., R.G. Weilbaecher, M. Walterscheid, et V. Lundblad. (2000). Identification of the single-strand telomeric DNA binding domain of the *Saccharomyces cerevisiae* Cdc13 protein. *Proc Natl Acad Sci U S A* **97**, 6457-62.

Hutchison, S., C. LeBel, M. Blanchette, et B. Chabot. (2002). Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *J Biol Chem* **277**, 29745-52.

Hwang, S.P. et R.S. Kucherlapati. (1983). Events preceding stable integration of SV40 genomes in a human cell line. *Somatic Cell Genet* **9**, 457-68.

Iervolino, A., G. Santilli, R. Trotta, C. Guerzoni, V. Cesi, A. Bergamaschi, et al. (2002). hnRNP A1 nucleocytoplasmic shuttling activity is required for normal myelopoiesis and BCR/ABL leukemogenesis. *Mol Cell Biol* **22**, 2255-66.

Ishikawa, F., M.J. Matunis, G. Dreyfuss, et T.R. Cech. (1993). Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)_n. *Mol Cell Biol* **13**, 4301-10.

Izaurralde, E., A. Jarmolowski, C. Beisel, I.W. Mattaj, G. Dreyfuss, et U. Fischer. (1997). A role for the M9 transport signal of hnRNP A1 in mRNA nuclear export. *J Cell Biol* **137**, 27-35.

- Jacob, N.K., K.E. Kirk, et C.M. Price. (2003). Generation of telomeric G strand overhangs involves both G and C strand cleavage. *Mol Cell* **11**, 1021-32.
- Jacob, N.K., R. Skopp, et C.M. Price. (2001). G-overhang dynamics at Tetrahymena telomeres. *EMBO J* **20**, 4299-308.
- Jahn, C.L. et L.A. Klobutcher. (2002). Genome remodeling in ciliated protozoa. *Annu Rev Microbiol* **56**, 489-520.
- James, P., J. Halladay, et E.A. Craig. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**, 1425-36.
- Jia, X., T. Weinert, et D. Lydall. (2004). Mec1 and Rad53 inhibit formation of single-stranded DNA at telomeres of *Saccharomyces cerevisiae* cdc13-1 mutants. *Genetics* **166**, 753-64.
- Jiang, Z.H., W.J. Zhang, Y. Rao, et J.Y. Wu. (1998). Regulation of Ich-1 pre-mRNA alternative splicing and apoptosis by mammalian splicing factors. *Proc Natl Acad Sci U S A* **95**, 9155-60.

- Johnson, R.E., G.K. Kovvali, L. Prakash, et S. Prakash. (1998). Role of yeast Rth1 nuclease and its homologs in mutation avoidance, DNA repair, and DNA replication. *Curr Genet* **34**, 21-9.
- Johnston, S.D., J.E. Lew, et J. Berman. (1999). Gbp1p, a protein with RNA recognition motifs, binds single-stranded telomeric DNA and changes its binding specificity upon dimerization. *Mol Cell Biol* **19**, 923-33.
- Jurica, M.S. et M.J. Moore. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* **12**, 5-14.
- Kajita, Y., J. Nakayama, M. Aizawa, et F. Ishikawa. (1995). The UUAG-specific RNA binding protein, heterogeneous nuclear ribonucleoprotein D0. Common modular structure and binding properties of the 2xRBD-Gly family. *J Biol Chem* **270**, 22167-75.
- Kanoh, J. et F. Ishikawa. (2003). Composition and conservation of the telomeric complex. *Cell Mol Life Sci* **60**, 2295-302.
- Kapoor, V. et W.G. Telford. (2004). Telomere length measurement by fluorescence in situ hybridization and flow cytometry. *Methods Mol Biol* **263**, 385-98.

- Karlseder, J., D. Broccoli, Y. Dai, S. Hardy, et T. de Lange. (1999). p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* **283**, 1321-5.
- Karlseder, J., A. Smogorzewska, et T. de Lange. (2002). Senescence induced by altered telomere state, not telomere loss. *Science* **295**, 2446-9.
- Kashima, T. et J.L. Manley. (2003). A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat Genet* **34**, 460-3.
- Kelleher, C., I. Kurth, et J. Lingner. (2005). Human protection of telomeres 1 (POT1) is a negative regulator of telomerase activity in vitro. *Mol Cell Biol* **25**, 808-18.
- Kelleher, C., M.T. Teixeira, K. Forstemann, et J. Lingner. (2002). Telomerase: biochemical considerations for enzyme and substrate. *Trends Biochem Sci* **27**, 572-9.
- Keys, B., V. Serra, G. Saretzki, et T. Von Zglinicki. (2004). Telomere shortening in human fibroblasts is not dependent on the size of the telomeric-3'-overhang. *Aging Cell* **3**, 103-9.
- Khan, S.A. (2005). Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid* **53**, 126-36.

Kim, N.W., M.A. Piatyszek, K.R. Prowse, C.B. Harley, M.D. West, P.L. Ho, et al.

(1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011-5.

Kipling, D. et H.J. Cooke. (1990). Hypervariable ultra-long telomeres in mice. *Nature* **347**, 400-2.

Kironmai, K.M. et K. Muniyappa. (1997). Alteration of telomeric sequences and senescence caused by mutations in RAD50 of *Saccharomyces cerevisiae*. *Genes Cells* **2**, 443-55.

Klobutcher, L.A., M.T. Swanton, P. Donini, et D.M. Prescott. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc Natl Acad Sci U S A* **78**, 3015-9.

Kohtz, J.D., S.F. Jamison, C.L. Will, P. Zuo, R. Luhrmann, M.A. Garcia-Blanco, et al. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**, 119-24.

- Krawczak, M., J. Reiss, et D.N. Cooper. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* **90**, 41-54.
- Krecic, A.M. et M.S. Swanson. (1999). hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* **11**, 363-71.
- Kumar, A. et S.H. Wilson. (1990). Studies of the strand-annealing activity of mammalian hnRNP complex protein A1. *Biochemistry* **29**, 10717-22.
- Labourier, E., M.D. Adams, et D.C. Rio. (2001). Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol Cell* **8**, 363-73.
- LaBranche, H., S. Dupuis, Y. Ben-David, M.R. Bani, R.J. Wellinger, et B. Chabot. (1998). Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat Genet* **19**, 199-202.
- Lallena, M.J., K.J. Chalmers, S. Llamazares, A.I. Lamond, et J. Valcarcel. (2002). Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* **109**, 285-96.

Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, et al. (2001).

Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Lareau, L.F., R.E. Green, R.S. Bhatnagar, et S.E. Brenner. (2004). The evolving roles of alternative splicing. *Curr Opin Struct Biol* **14**, 273-82.

Larrivée, M., C. LeBel, et R.J. Wellinger. (2004). The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. *Genes Dev* **18**, 1391-6.

Laski, F.A., D.C. Rio, et G.M. Rubin. (1986). Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. *Cell* **44**, 7-19.

Le, S., J.K. Moore, J.E. Haber, et C.W. Greider. (1999). RAD50 and RAD51 define two pathways that collaborate to maintain telomeres in the absence of telomerase. *Genetics* **152**, 143-52.

LeBel, C. et R.J. Wellinger. (2004). The structures at the termini of chromosomes: what is there hidden under the cap ?. *Med Sci (Paris)* **20**, 207-12.

- Lei, M., P. Baumann, et T.R. Cech. (2002). Cooperative binding of single-stranded telomeric DNA by the Pot1 protein of *Schizosaccharomyces pombe*. *Biochemistry* **41**, 14560-8.
- Lei, M., E.R. Podell, P. Baumann, et T.R. Cech. (2003). DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. *Nature* **426**, 198-203.
- Lendvay, T.S., D.K. Morris, J. Sah, B. Balasubramanian, et V. Lundblad. (1996). Senescence mutants of *Saccharomyces cerevisiae* with a defect in telomere replication identify three additional EST genes. *Genetics* **144**, 1399-412.
- LeStourgeon, W.M., A.L. Beyer, M.E. Christensen, B.W. Walker, S.M. Poupore, et L.P. Daniels. (1978). The packaging proteins of core hnRNP particles and the maintenance of proliferative cell states. *Cold Spring Harb Symp Quant Biol* **42 Pt 2**, 885-98.
- Levy, T., I. AgoulNIK, E.N. Atkinson, X.W. Tong, H.M. Gause, A. Hasenburger, et al. (1998). Telomere length in human white blood cells remains constant with age and is shorter in breast cancer patients. *Anticancer Res* **18**, 1345-9.

- Li, G.Z., M.S. Eller, R. Firoozabadi, et B.A. Gilchrest. (2003). Evidence that exposure of the telomere 3' overhang sequence induces senescence. *Proc Natl Acad Sci U S A* **100**, 527-31.
- Libri, D., A. Piseri, et M.Y. Fiszman. (1991). Tissue-specific splicing in vivo of the beta-tropomyosin gene: dependence on an RNA secondary structure. *Science* **252**, 1842-5.
- Lim, K., J.X. Ho, K. Keeling, G.L. Gilliland, X. Ji, F. Ruker, et al. (1994). Three-dimensional structure of *Schistosoma japonicum* glutathione S-transferase fused with a six-amino acid conserved neutralizing epitope of gp41 from HIV. *Protein Sci* **3**, 2233-44.
- Lin, J.J. et V.A. Zakian. (1996). The *Saccharomyces* CDC13 protein is a single-strand TG1-3 telomeric DNA-binding protein in vitro that affects telomere behavior in vivo. *Proc Natl Acad Sci U S A* **93**, 13760-5.
- Lin, Y.C., C.L. Hsu, J.W. Shih, et J.J. Lin. (2001). Specific binding of single-stranded telomeric DNA by Cdc13p of *Saccharomyces cerevisiae*. *J Biol Chem* **276**, 24588-93.

- Lindstrom, U.M., R.A. Chandrasekaran, L. Orbai, S.A. Helquist, G.P. Miller, E. Oroudjev, et al. (2002). Artificial human telomeres from DNA nanocircle templates. *Proc Natl Acad Sci U S A* **99**, 15953-8.
- Lingner, J. et T.R. Cech. (1996). Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3' overhang. *Proc Natl Acad Sci U S A* **93**, 10712-7.
- Lingner, J., T.R. Cech, T.R. Hughes, et V. Lundblad. (1997a). Three Ever Shorter Telomere (EST) genes are dispensable for in vitro yeast telomerase activity. *Proc Natl Acad Sci U S A* **94**, 11190-5.
- Lingner, J., J.P. Cooper, et T.R. Cech. (1995). Telomerase and DNA end replication: no longer a lagging strand problem? *Science* **269**, 1533-4.
- Lingner, J., T.R. Hughes, A. Shevchenko, M. Mann, V. Lundblad, et T.R. Cech. (1997b). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561-7.
- Liu, D., M.S. O'Connor, J. Qin, et Z. Songyang. (2004a). Telosome -- A mammalian telomere associated complex formed by multiple telomeric proteins. *J Biol Chem*.

- Liu, D., A. Safari, M.S. O'Connor, D.W. Chan, A. Laegeler, J. Qin, et al. (2004b). PTP interacts with POT1 and regulates its localization to telomeres. *Nat Cell Biol* **6**, 673-80.
- Liu, Q. et G. Dreyfuss. (1995). In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol Cell Biol* **15**, 2800-8.
- Loayza, D. et T. De Lange. (2003). POT1 as a terminal transducer of TRF1 telomere length control. *Nature* **424**, 1013-8.
- Loayza, D., H. Parsons, J. Donigian, K. Hoke, et T. De Lange. (2004). DNA binding features of human POT1: A nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites. *J Biol Chem.* **279**, 13241-8.
- Lombard, D.B. et L. Guarente. (2000). Nijmegen breakage syndrome disease protein and MRE11 at PML nuclear bodies and meiotic telomeres. *Cancer Res* **60**, 2331-4.
- Longman, D., I.L. Johnstone, et J.F. Cáceres. (2000). Functional characterization of SR and SR-related genes in *Caenorhabditis elegans*. *EMBO J* **19**, 1625-37.

Lopez, P.J. et B. Seraphin. (2000). YIDB: the Yeast Intron DataBase. *Nucleic Acids Res* **28**, 85-6.

Louis, E.J. (1995). The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* **11**, 1553-73.

Louis, E.J. et J.E. Haber. (1990). Mitotic recombination among subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* **124**, 547-59.

-. (1992). The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* **131**, 559-74.

Louis, E.J., E.S. Naumova, A. Lee, G. Naumov, et J.E. Haber. (1994). The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* **136**, 789-802.

Lue, N.F. (2004). Adding to the ends: what makes telomerase processive and how important is it? *Bioessays* **26**, 955-62.

Lundblad, V. et E.H. Blackburn. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell* **73**, 347-60.

- Lundblad, V. et J.W. Szostak. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**, 633-43.
- Lustig, A.J. (2001). Cdc13 subcomplexes regulate multiple telomere functions. *Nat Struct Biol* **8**, 297-9.
- . (2003). Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. *Nat Rev Genet* **4**, 916-23.
- Lydall, D. et T. Weinert. (1995). Yeast checkpoint genes in DNA damage processing: implications for repair and arrest. *Science* **270**, 1488-91.
- Lynch, K.W. et T. Maniatis. (1996). Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes Dev* **10**, 2089-101.
- MacKenzie, K.L., S. Franco, C. May, M. Sadelain, et M.A. Moore. (2000). Mass cultured human fibroblasts overexpressing hTERT encounter a growth crisis following an extended period of proliferation. *Exp Cell Res* **259**, 336-50.

Makarov, V.L., Y. Hirose, et J.P. Langmore. (1997). Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening.

Cell **88**, 657-66.

Maniatis, T. et R. Reed. (2002). An extensive network of coupling among gene expression machines. *Nature* **416**, 499-506.

Maniatis, T. et B. Tasic. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236-43.

Manley, J.L. et R. Tacke. (1996). SR proteins and splicing control. *Genes Dev* **10**, 1569-79.

Marcand, S., V. Brevet, et E. Gilson. (1999). Progressive cis-inhibition of telomerase upon telomere elongation. *EMBO J* **18**, 3509-19.

Marcand, S., V. Brevet, C. Mann, et E. Gilson. (2000). Cell cycle restriction of telomere elongation. *Curr Biol* **10**, 487-90.

Martens, U.M., E.A. Chavez, S.S. Poon, C. Schmoor, et P.M. Lansdorp. (2000). Accumulation of short telomeres in human fibroblasts prior to replicative senescence. *Exp Cell Res* **256**, 291-9.

Masutomi, K., E.Y. Yu, S. Khurts, I. Ben-Porath, J.L. Currier, G.B. Metz, et al. (2003).

Telomerase maintains telomere structure in normal human cells. *Cell* **114**, 241-53.

Matunis, E.L., M.J. Matunis, et G. Dreyfuss. (1993). Association of individual hnRNP proteins and snRNPs with nascent transcripts. *J Cell Biol* **121**, 219-28.

Mayeda, A., D.M. Helfman, et A.R. Krainer. (1993). Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol Cell Biol* **13**, 2993-3001.

Mayeda, A. et A.R. Krainer. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**, 365-75.

Mayeda, A., S.H. Munroe, J.F. Cáceres, et A.R. Krainer. (1994). Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J* **13**, 5483-95.

Mayeda, A., S.H. Munroe, R.M. Xu, et A.R. Krainer. (1998). Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. *RNA* **4**, 1111-23.

Mayeda, A., G.R. Screaton, S.D. Chandler, X.D. Fu, et A.R. Krainer. (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA

- recognition motifs and composite pre-mRNA exonic elements. *Mol Cell Biol* **19**, 1853-63.
- McClintock, B. (1939). The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proc Natl Acad Sci U S A* **25**, 405-416.
- . (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234-282.
- McElligott, R. et R.J. Wellinger. (1997). The terminal DNA structure of mammalian chromosomes. *EMBO J* **16**, 3705-14.
- McKay, S.J. et H. Cooke. (1992a). hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGGn. *Nucleic Acids Res* **20**, 6461-4.
- . (1992b). A protein which specifically binds to single stranded TTAGGGn repeats. *Nucleic Acids Res* **20**, 1387-91.
- McTigue, M.A., D.R. Williams, et J.A. Tainer. (1995). Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug praziquantel. *J Mol Biol* **246**, 21-7.

- Meier, B., L. Driller, S. Jaklin, et H.M. Feldmann. (2001). New function of CDC13 in positive telomere length regulation. *Mol Cell Biol* **21**, 4233-45.
- Mermoud, J.E., P.T. Cohen, et A.I. Lamond. (1994). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J* **13**, 5679-88.
- Merrill, B.M., M.B. Lopresti, K.L. Stone, et K.R. Williams. (1987). Amino acid sequence of UP1, an hnRNP-derived single-stranded nucleic acid binding protein from calf thymus. *Int J Pept Protein Res* **29**, 21-39.
- Merrill, B.M., K.L. Stone, F. Cobianchi, S.H. Wilson, et K.R. Williams. (1988). Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. *J Biol Chem* **263**, 3307-13.
- Meyerson, M., C.M. Counter, E.N. Eaton, L.W. Ellisen, P. Steiner, S.D. Caddle, et al. (1997). hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. *Cell* **90**, 785-95.

Meyne, J., R.L. Ratliff, et R.K. Moyzis. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc Natl Acad Sci U S A* **86**, 7049-53.

Michael, W.M., M. Choi, et G. Dreyfuss. (1995). A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* **83**, 415-22.

Mieczkowski, P.A., M.U. Fikus, et Z. Ciesla. (1997). Characterization of a novel DNA damage-inducible gene of *Saccharomyces cerevisiae*, DIN7, which is a structural homolog of the RAD2 and RAD27 DNA repair genes. *Mol Gen Genet* **253**, 655-65.

Mitton-Fry, R.M., E.M. Anderson, T.R. Hughes, V. Lundblad, et D.S. Wuttke. (2002). Conserved structure for single-stranded telomeric DNA recognition. *Science* **296**, 145-7.

Mitton-Fry, R.M., E.M. Anderson, D.L. Theobald, L.W. Glustrom, et D.S. Wuttke. (2004). Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13. *J Mol Biol* **338**, 241-55.

Modrek, B. et C.J. Lee. (2003). Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss.

Nat Genet **34**, 177-80.

Molenaar, C., K. Wiesmeijer, N.P. Verwoerd, S. Khazen, R. Eils, H.J. Tanke, et al.

(2003). Visualizing telomere dynamics in living mammalian cells using PNA probes. *EMBO J* **22**, 6631-41.

Moore, P.D., J.R. Simon, L.J. Wallace, et T.Y. Chow. (1993). In-vitro recombination in

rad and rnc mutants of *Saccharomyces cerevisiae*. *Curr Genet* **23**, 1-8.

Moreau, S., J.R. Ferguson, et L.S. Symington. (1999). The nuclease activity of Mre11 is

required for meiosis but not for mating type switching, end joining, or telomere maintenance. *Mol Cell Biol* **19**, 556-66.

Morin, G.B. (1989). The human telomere terminal transferase enzyme is a

ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* **59**, 521-9.

Moyzis, R.K. (1991). The human telomere. *Sci Am* **265**, 48-55.

Moyzis, R.K., J.M. Buckingham, L.S. Cram, M. Dani, L.L. Deaven, M.D. Jones, et al.

(1988). A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A* **85**, 6622-6.

Muggleton-Harris, A.L. et L. Hayflick. (1976). Cellular aging studied by the reconstruction of replicating cells from nuclei and cytoplasms isolated from normal human diploid cells. *Exp Cell Res* **103**, 321-30.

Muller, H.J. (1938). The remaking of chromosomes. *The Collecting Net* **13**, 181-95, 198.

Munoz-Jordan, J.L., G.A. Cross, T. de Lange, et J.D. Griffith. (2001). t-loops at trypanosome telomeres. *EMBO J* **20**, 579-88.

Munroe, S.H. et X.F. Dong. (1992). Heterogeneous nuclear ribonucleoprotein A1 catalyzes RNA.RNA annealing. *Proc Natl Acad Sci U S A* **89**, 895-9.

Murnane, J.P., L. Sabatier, B.A. Marder, et W.F. Morgan. (1994). Telomere dynamics in an immortal human cell line. *EMBO J* **13**, 4953-62.

Muro, A.F., A. Iaconcig, et F.E. Baralle. (1998). Regulation of the fibronectin EDA exon alternative splicing. Cooperative role of the exonic enhancer element and the 5' splicing site. *FEBS Lett* **437**, 137-41.

- Murti, K.G. et D.M. Prescott. (1999). Telomeres of polytene chromosomes in a ciliated protozoan terminate in duplex DNA loops. *Proc Natl Acad Sci U S A* **96**, 14436-9.
- Murzin, A.G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J* **12**, 861-7.
- Myers, J.C. et Y. Shamoo. (2004). Human UP1 as a model for understanding purine recognition in the family of proteins containing the RNA recognition motif (RRM). *J Mol Biol* **342**, 743-56.
- Nakamura, T.M., G.B. Morin, K.B. Chapman, S.L. Weinrich, W.H. Andrews, J. Lingner, et al. (1997). Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**, 955-9.
- Nasim, F.U., S. Hutchison, M. Cordeau, et B. Chabot. (2002). High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA* **8**, 1078-89.

- Natarajan, S., C. Groff-Vindman, et M.J. McEachern. (2003). Factors influencing the recombinational expansion and spread of telomeric tandem arrays in *Kluyveromyces lactis*. *Eukaryot Cell* **2**, 1115-27.
- Natarajan, S. et M.J. McEachern. (2002). Recombinational telomere elongation promoted by DNA circles. *Mol Cell Biol* **22**, 4512-21.
- Nemoto, T., Y. Ohara-Nemoto, S. Shimazaki, et M. Ota. (1994). Dimerization characteristics of the DNA- and steroid-binding domains of the androgen receptor. *J Steroid Biochem Mol Biol* **50**, 225-33.
- Niedziela-Majka, A., G. Rymarczyk, M. Kochman, et A. Ozyhar. (1998). GST-Induced dimerization of DNA-binding domains alters characteristics of their interaction with DNA. *Protein Expr Purif* **14**, 208-20.
- Nikitina, T. et C.L. Woodcock. (2004). Closed chromatin loops at the ends of chromosomes. *J Cell Biol* **166**, 161-5.
- Nilsen, T.W. (2002). The spliceosome: no assembly required? *Mol Cell* **9**, 8-9.
- . (2003). The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* **25**, 1147-9.

Novick, R.P. (1998). Contrasting lifestyles of rolling-circle phages and plasmids. *Trends Biochem Sci* **23**, 434-8.

Nugent, C.I., G. Bosco, L.O. Ross, S.K. Evans, A.P. Salinger, J.K. Moore, et al. (1998).

Telomere maintenance is dependent on activities required for end repair of double-strand breaks. *Curr Biol* **8**, 657-60.

Nugent, C.I., T.R. Hughes, N.F. Lue, et V. Lundblad. (1996). Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science* **274**, 249-52.

Ogino, H., K. Nakabayashi, M. Suzuki, E. Takahashi, M. Fujii, T. Suzuki, et al. (1998).

Release of telomeric DNA from chromosomes in immortal human cells lacking telomerase activity. *Biochem Biophys Res Commun* **248**, 223-7.

Olovnikov, A.M. (1973). A theory of marginotomy. *J Theor Biol* **41**, 181-190.

Paques, F. et J.E. Haber. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **63**, 349-404.

Patel, A.A. et J.A. Steitz. (2003). Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* **4**, 960-70.

Patry, C., L. Bouchard, P. Labrecque, D. Gendron, B. Lemieux, J. Toutant, et al. (2003).

Small interfering RNA-mediated reduction in heterogeneous nuclear ribonucleoparticle A1/A2 proteins induces apoptosis in human cancer cells but not in normal mortal cell lines. *Cancer Res* **63**, 7679-88.

Patry, C., B. Lemieux, R.J. Wellinger, et B. Chabot. (2004). Targeting heterogeneous nuclear ribonucleoparticle A1 and A2 proteins by RNA interference promotes cell death in transformed but not in normal mouse cell lines. *Mol Cancer Ther* **3**, 1193-9.

Paull, T.T. et M. Gellert. (1999). Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes Dev* **13**, 1276-88.

Peng, Y., I.S. Mian, et N.F. Lue. (2001). Analysis of telomerase processivity: mechanistic similarity to HIV-1 reverse transcriptase and role in telomere maintenance. *Mol Cell* **7**, 1201-11.

Pennisi, E. (2000). Human Genome Project. And the gene number is...? *Science* **288**, 1146-7.

- Pennock, E., K. Buckley, et V. Lundblad. (2001). Cdc13 Delivers Separate Complexes to the Telomere for End Protection and Replication. *Cell* **104**, 387-96.
- Perrem, K., L.M. Colgin, A.A. Neumann, T.R. Yeager, et R.R. Reddel. (2001). Coexistence of alternative lengthening of telomeres and telomerase in hTERT-transfected GM847 cells. *Mol Cell Biol* **21**, 3862-75.
- Petersen, S., G. Saretzki, et T. von Zglinicki. (1998). Preferential accumulation of single-stranded regions in telomeres of human fibroblasts. *Exp Cell Res* **239**, 152-60.
- Petrov, A.V., S.S. Dokudovskaya, K.A. Sokolov, O.I. Lavrik, A. Favre, O.A. Dontsova, et al. (1998). Telomerase from *Saccharomyces cerevisiae* contains several protein subunits and may have different activities depending on the protein content. *FEBS Lett* **436**, 35-40.
- Philips, A.V. et T.A. Cooper. (2000). RNA processing and human disease. *Cell Mol Life Sci* **57**, 235-49.
- Pinol-Roma, S., S.A. Adam, Y.D. Choi, et G. Dreyfuss. (1989). Ultraviolet-induced cross-linking of RNA to proteins in vivo. *Methods Enzymol* **180**, 410-8.

- Pinol-Roma, S., Y.D. Choi, M.J. Matunis, et G. Dreyfuss. (1988). Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev* **2**, 215-27.
- Pluta, A.F., B.P. Kaine, et B.B. Spear. (1982). The terminal organization of macronuclear DNA in *Oxytricha fallax*. *Nucleic Acids Res* **10**, 8145-54.
- Prasad, J., K. Colwill, T. Pawson, et J.L. Manley. (1999). The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Mol Cell Biol* **19**, 6991-7000.
- Price, C.M. (1990). Telomere structure in *Euplotes crassus*: characterization of DNA-protein interactions and isolation of a telomere-binding protein. *Mol Cell Biol* **10**, 3421-31.
- Price, C.M. et T.R. Cech. (1987). Telomeric DNA-protein interactions of *Oxytricha* macronuclear DNA. *Genes Dev* **1**, 783-93.
- Prowse, K.R., A.A. Avilion, et C.W. Greider. (1993). Identification of a nonprocessive telomerase activity from mouse cells. *Proc Natl Acad Sci U S A* **90**, 1493-7.

- Prowse, K.R. et C.W. Greider. (1995). Developmental and tissue-specific regulation of mouse telomerase and telomere length. *Proc Natl Acad Sci U S A* **92**, 4818-22.
- Pryde, F.E. et E.J. Louis. (1997). *Saccharomyces cerevisiae* telomeres. A review. *Biochemistry (Mosc)* **62**, 1232-41.
- . (1999). Limitations of silencing at native yeast telomeres. *EMBO J* **18**, 2538-50.
- Qi, H. et V.A. Zakian. (2000). The *Saccharomyces* telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein. *Genes Dev* **14**, 1777-88.
- Raghunathan, S., A.G. Kozlov, T.M. Lohman, et G. Waksman. (2000). Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nat Struct Biol* **7**, 648-52.
- Rappsilber, J., U. Ryder, A.I. Lamond, et M. Mann. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**, 1231-45.
- Raymond, W.E. et N. Kleckner. (1993). RAD50 protein of *S.cerevisiae* exhibits ATP-dependent DNA binding. *Nucleic Acids Res* **21**, 3851-6.
- Reddel, R.R. (2003). Alternative lengthening of telomeres, telomerase, and cancer. *Cancer Lett* **194**, 155-62.

- Reddel, R.R. et T.M. Bryan. (2003). Alternative lengthening of telomeres: dangerous road less travelled. *Lancet* **361**, 1840-1.
- Reddel, R.R., T.M. Bryan, et J.P. Murnane. (1997). Immortalized cells with no detectable telomerase activity. A review. *Biochemistry (Mosc)* **62**, 1254-62.
- Richards, E.J. et F.M. Ausubel. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* **53**, 127-36.
- Riha, K., J. Fajkus, J. Siroky, et B. Vyskot. (1998). Developmental control of telomere lengths and telomerase activity in plants. *Plant Cell* **10**, 1691-8.
- Riha, K., T.D. McKnight, J. Fajkus, B. Vyskot, et D.E. Shippen. (2000). Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. *Plant J* **23**, 633-41.
- Ritchie, K.B. et T.D. Petes. (2000). The Mre11p/Rad50p/Xrs2p complex and the Tel1p function in a single pathway for telomere maintenance in yeast. *Genetics* **155**, 475-9.

Riva, S., C. Morandi, P. Tsoulfas, M. Pandolfo, G. Biamonti, B. Merrill, et al. (1986).

Mammalian single-stranded DNA binding protein UP I is derived from the hnRNP core protein A1. *EMBO J* **5**, 2267-73.

Robberson, B.L., G.J. Côté, et S.M. Berget. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**, 84-94.

Roest Crolius, H., O. Jaillon, A. Bernot, C. Dasilva, L. Bouneau, C. Fischer, et al. (2000). Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat Genet* **25**, 235-8.

Rose, M.D., F. Winston, et P. Hieter. (1990). *Methods in Yeast Genetics: A Laboratory Course Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sacco-Bubulya, P. et D.L. Spector. (2002). Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing. *J Cell Biol* **156**, 425-36.

Sambrook, J., E.F. Fritsch, et T. Maniatis. (1989). *Molecular Cloning: A Laboratory Manual, Second Edition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanchez-Velar, N., E.B. Udofia, Z. Yu, et M.L. Zapp. (2004). hRIP, a cellular cofactor for Rev function, promotes release of HIV RNAs from the perinuclear region.

Genes Dev **18**, 23-34.

Sanford, J.R. et J.P. Bruzik. (1999). Developmental regulation of SR protein phosphorylation and activity. *Genes Dev* **13**, 1513-8.

-. (2001). Regulation of SR protein localization during development. *Proc Natl Acad Sci U S A* **98**, 10184-9.

Sarig, G., P. Weisman-Shomer, R. Erlitzki, et M. Fry. (1997). Purification and characterization of qTBP42, a new single-stranded and quadruplex telomeric DNA-binding protein from rat hepatocytes. *J Biol Chem* **272**, 4474-82.

Satoh, H., H. Kamma, H. Ishikawa, H. Horiguchi, M. Fujiwara, Y.T. Yamashita, et al. (2000). Expression of hnRNP A2/B1 proteins in human cancer cell lines. *Int J Oncol* **16**, 763-7.

Schmucker, D., J.C. Clemens, H. Shu, C.A. Worby, J. Xiao, M. Muda, et al. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671-84.

Schramke, V., P. Luciano, V. Brevet, S. Guillot, Y. Corda, M.P. Longhese, et al. (2004).

RPA regulates telomerase action by providing Est1p access to chromosome ends.

Nat Genet **36**, 46-54.

Seto, A.G., K. Umansky, Y. Tzfati, A.J. Zaug, E.H. Blackburn, et T.R. Cech. (2003). A

template-proximal RNA paired element contributes to *Saccharomyces cerevisiae*

telomerase activity. *RNA* **9**, 1323-32.

Sfeir, A.J., W. Chai, J.W. Shay, et W.E. Wright. (2005). Telomere-end processing the

terminal nucleotides of human chromosomes. *Mol Cell* **18**, 131-8.

Shammas, M.A., X. Liu, G. Gavory, K.D. Raney, S. Balasubramanian, et R.J. Shmookler

Reis. (2004). Targeting the single-strand G-rich overhang of telomeres with PNA

inhibits cell growth and induces apoptosis of human immortal cells. *Exp Cell Res*

295, 204-14.

Shamoo, Y., N. Abdul-Manan, A.M. Patten, J.K. Crawford, M.C. Pellegrini, et K.R.

Williams. (1994). Both RNA-binding domains in heterogenous nuclear

ribonucleoprotein A1 contribute toward single-stranded-RNA binding.

Biochemistry **33**, 8272-81.

- Shampay, J., J.W. Szostak, et E.H. Blackburn. (1984). DNA sequences of telomeres maintained in yeast. *Nature* **310**, 154-7.
- Sharp, P.A. (1994). Split genes and RNA splicing. *Cell* **77**, 805-15.
- Sheng, H., Z. Hou, T. Schierer, D.L. Dobbs, et E. Henderson. (1995). Identification and characterization of a putative telomere end-binding protein from *Tetrahymena thermophila*. *Mol Cell Biol* **15**, 1144-53.
- Shippen-Lentz, D. et E.H. Blackburn. (1989). Telomere terminal transferase activity from *Euplotes crassus* adds large numbers of TTTTGGGG repeats onto telomeric primers. *Mol Cell Biol* **9**, 2761-4.
- Siebel, C.W., A. Admon, et D.C. Rio. (1995). Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes Dev* **9**, 269-83.
- Siebel, C.W., L.D. Fresco, et D.C. Rio. (1992). The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. *Genes Dev* **6**, 1386-401.
- Simard, M.J. et B. Chabot. (2000). Control of hnRNP A1 alternative splicing: an intron element represses use of the common 3' splice site. *Mol Cell Biol* **20**, 7353-62.

- . (2002). SRp30c is a repressor of 3' splice site utilization. *Mol Cell Biol* **22**, 4001-10.
- Singer, M.S. et D.E. Gottschling. (1994). TLC1: template RNA component of *Saccharomyces cerevisiae* telomerase. *Science* **266**, 404-9.
- Smith, C.D., D.L. Smith, J.L. DeRisi, et E.H. Blackburn. (2003). Telomeric protein distributions and remodeling through the cell cycle in *Saccharomyces cerevisiae*. *Mol Biol Cell* **14**, 556-70.
- Smogorzewska, A. et T. de Lange. (2004). Regulation of telomerase by telomeric proteins. *Annu Rev Biochem* **73**, 177-208.
- Snead, D.R., B. Perunovic, N. Cullen, M. Needham, D.P. Dhillon, H. Satoh, et al. (2003). hnRNP B1 expression in benign and malignant lung disease. *J Pathol* **200**, 88-94.
- Solnick, D. (1985). Alternative splicing caused by RNA secondary structure. *Cell* **43**, 667-76.
- Solnick, D. et S.I. Lee. (1987). Amount of RNA secondary structure required to induce an alternative splice. *Mol Cell Biol* **7**, 3194-8.
- Spingola, M. et M. Ares, Jr. (2000). A yeast intronic splicing enhancer and Nam8p are required for Mer1p-activated splicing. *Mol Cell* **6**, 329-38.

- Staley, J.P. et C. Guthrie. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-26.
- Stamm, S.(2002). Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum Mol Genet* **11**, 2409-16.
- Stamm, S., S. Ben-Ari, I. Rafalska, Y. Tang, Z. Zhang, D. Toiber, et al. (2005). Function of alternative splicing. *Gene* **344C**, 1-20.
- Stansel, R.M., T. de Lange, et J.D. Griffith. (2001). T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang. *EMBO J* **20**, 5532-40.
- Starling, J.A., J. Maule, N.D. Hastie, et R.C. Allshire. (1990). Extensive telomere repeat arrays in mouse are hypervariable. *Nucleic Acids Res* **18**, 6881-8.
- Steiner, B.R., K. Hidaka, et B. Futcher. (1996). Association of the Est1 protein with telomerase activity in yeast. *Proc Natl Acad Sci U S A* **93**, 2817-21.
- Steinert, S., J.W. Shay, et W.E. Wright. (2004). Modification of subtelomeric DNA. *Mol Cell Biol* **24**, 4571-80.

Stevens, S.W., D.E. Ryan, H.Y. Ge, R.E. Moore, M.K. Young, T.D. Lee, et al. (2002).

Composition and functional characterization of the yeast spliceosomal pentanucleotide small nuclear ribonucleoprotein (snRNP). *Mol Cell* **9**, 31-44.

Stewart, S.A., I. Ben-Porath, V.J. Carey, B.F. O'Connor, W.C. Hahn, et R.A. Weinberg.

(2003). Erosion of the telomeric single-strand overhang at replicative senescence.

Nat Genet **33**, 492-6.

Stewart, S.A. et R.A. Weinberg. (2002). Senescence: does it all happen at the ends?

Oncogene **21**, 627-30.

Sueoka, E., Y. Goto, N. Sueoka, Y. Kai, T. Kozu, et H. Fujiki. (1999). Heterogeneous

nuclear ribonucleoprotein B1 as a new marker of early detection for human lung

cancers. *Cancer Res* **59**, 1404-7.

Sugawara, N. et J.W. Szostak. (1986). Telomeres of *Schizosaccharomyces pombe*. *Yeast*

(Suppl.) **1**, 373.

Swanson, M.S. et G. Dreyfuss. (1988). RNA binding specificity of hnRNP proteins: a

subset bind to the 3' end of introns. *EMBO J* **7**, 3519-29.

- Tacke, R. et J.L. Manley. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J* **14**, 3540-51.
- . (1999). Determinants of SR protein specificity. *Curr Opin Cell Biol* **11**, 358-62.
- Taggart, A.K., S.C. Teng, et V.A. Zakian. (2002). Est1p as a cell cycle-regulated activator of telomere-bound telomerase. *Science* **297**, 1023-6.
- Takai, H., A. Smogorzewska, et T. de Lange. (2003). DNA damage foci at dysfunctional telomeres. *Curr Biol* **13**, 1549-56.
- Takata, H., Y. Tanaka, et A. Matsuura. (2005). Late S phase-specific recruitment of Mre11 complex triggers hierarchical assembly of telomere replication proteins in *Saccharomyces cerevisiae*. *Mol Cell* **17**, 573-83.
- Tarn, W.Y. et J.A. Steitz. (1997). Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci* **22**, 132-7.
- Teng, S.C., J. Chang, B. McCowan, et V.A. Zakian. (2000). Telomerase-independent lengthening of yeast telomeres occurs by an abrupt Rad50p-dependent, Rif-inhibited recombinational process. *Mol Cell* **6**, 947-952.

- Teng, S.C. et V.A. Zakian. (1999). Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in *Saccharomyces cerevisiae*. *Mol Cell Biol* **19**, 8083-93.
- Theobald, D.L., R.B. Cervantes, V. Lundblad, et D.S. Wuttke. (2003). Homology among telomeric end-protection proteins. *Structure (Camb)* **11**, 1049-50.
- Theobald, D.L. et S.C. Schultz. (2003). Nucleotide shuffling and ssDNA recognition in *Oxytricha nova* telomere end-binding protein complexes. *EMBO J* **22**, 4314-24.
- Theobald, D.L. et D.S. Wuttke. (2004). Prediction of multiple tandem OB-fold domains in telomere end-binding proteins Pot1 and Cdc13. *Structure (Camb)* **12**, 1877-9.
- Tian, M. et T. Maniatis. (1993). A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell* **74**, 105-14.
- Timmer, T., P. Terpstra, A. van den Berg, P.M. Veldhuis, A. Ter Elst, A.Y. van der Veen, et al. (1999a). An evolutionary rearrangement of the Xp11.3-11.23 region in 3p21.3, a region frequently deleted in a variety of cancers. *Genomics* **60**, 238-40.

Timmer, T., P. Terpstra, A. van den Berg, P.M. Veldhuis, A. Ter Elst, G. Voutsinas, et al.

(1999b). A comparison of genomic structures and expression patterns of two closely related flanking genes in a critical lung cancer region at 3p21.3. *Eur J Hum Genet* **7**, 478-86.

Tishkoff, D.X., A.L. Boerger, P. Bertrand, N. Filosi, G.M. Gaida, M.F. Kane, et al.

(1997). Identification and characterization of *Saccharomyces cerevisiae* EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proc Natl Acad Sci U S A* **94**, 7487-92.

Tokutake, Y., T. Matsumoto, T. Watanabe, S. Maeda, H. Tahara, S. Sakamoto, et al.

(1998). Extra-chromosomal telomere repeat DNA in telomerase-negative immortalized cell lines. *Biochem Biophys Res Commun* **247**, 765-72.

Tomaska, L., M.J. McEachern, et J. Nosek. (2004a). Alternatives to telomerase: keeping

linear chromosomes via telomeric circles. *FEBS Lett* **567**, 142-6.

Tomaska, L., J. Nosek, A.M. Makhov, A. Pastorakova, et J.D. Griffith. (2000).

Extragenomic double-stranded DNA circles in yeast with linear mitochondrial

genomes: potential involvement in telomere maintenance. *Nucleic Acids Res* **28**, 4479-87.

Tomaska, L., S. Willcox, J. Slezakova, J. Nosek, et J.D. Griffith. (2004b). Taz1 binding to a fission yeast model telomere: formation of telomeric loops and higher order structures. *J Biol Chem* **279**, 50764-72.

Tomita, K., T. Kibe, H.Y. Kang, Y.S. Seo, M. Uritani, T. Ushimaru, et al. (2004). Fission yeast Dna2 is required for generation of the telomeric single-strand overhang. *Mol Cell Biol* **24**, 9557-67.

Tsukamoto, Y., J. Kato, et H. Ikeda. (1996). Hdf1, a yeast Ku-protein homologue, is involved in illegitimate recombination, but not in homologous recombination. *Nucleic Acids Res* **24**, 2067-72.

Tsukamoto, Y., A.K. Taggart, et V.A. Zakian. (2001). The role of the Mre11-Rad50-Xrs2 complex in telomerase-mediated lengthening of *Saccharomyces cerevisiae* telomeres. *Curr Biol* **11**, 1328-35.

- Underwood, D.H., C. Carroll, et M.J. McEachern. (2004). Genetic dissection of the *Kluyveromyces lactis* telomere and evidence for telomere capping defects in TER1 mutants with long telomeres. *Eukaryot Cell* **3**, 369-84.
- Usui, T., T. Ohta, H. Oshiumi, J. Tomizawa, H. Ogawa, et T. Ogawa. (1998). Complex formation and functional versatility of Mre11 of budding yeast in recombination. *Cell* **95**, 705-16.
- Valcarcel, J., R. Singh, P.D. Zamore, et M.R. Green. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* **362**, 171-5.
- Vallano, M.L., B. Lambolez, E. Audinat, et J. Rossier. (1996). Neuronal activity differentially regulates NMDA receptor subunit expression in cerebellar granule cells. *J Neurosci* **16**, 631-9.
- Van der Ploeg, L.H., A.Y. Liu, et P. Borst. (1984). Structure of the growing telomeres of Trypanosomes. *Cell* **36**, 459-68.
- van Steensel, B., A. Smogorzewska, et T. de Lange. (1998). TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401-13.

- Veldman, T., K.T. Etheridge, et C.M. Counter. (2004). Loss of hPot1 function leads to telomere instability and a cut-like phenotype. *Curr Biol* **14**, 2264-70.
- Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, et al. (2001). The sequence of the human genome. *Science* **291**, 1304-51.
- Virta-Pearlman, V., D.K. Morris, et V. Lundblad. (1996). Est1 has the properties of a single-stranded telomere end-binding protein. *Genes Dev* **10**, 3094-104.
- Von Zglinicki, T. (2003). Replicative senescence and the art of counting. *Exp Gerontol* **38**, 1259-64.
- Walker, J., P. Crowley, A.D. Moreman, et J. Barrett.(1993). Biochemical properties of cloned glutathione S-transferases from *Schistosoma mansoni* and *Schistosoma japonicum*. *Mol Biochem Parasitol* **61**, 255-64.
- Walmsley, R.W., C.S. Chan, B.K. Tye, et T.D. Petes. (1984). Unusual DNA sequences associated with the ends of yeast chromosomes. *Nature* **310**, 157-60.
- Wang, H.Y., X. Xu, J.H. Ding, J.R. Bermingham, Jr., et X.D. Fu. (2001). SC35 plays a role in T cell development and alternative splicing of CD45. *Mol Cell* **7**, 331-42.

- Wang, J., S.H. Xiao, et J.L. Manley. (1998). Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing. *Genes Dev* **12**, 2222-33.
- Wang, M.J., Y.C. Lin, T.L. Pang, J.M. Lee, C.C. Chou, et J.J. Lin. (2000). Telomere-binding and Stn1p-interacting activities are required for the essential function of *Saccharomyces cerevisiae* Cdc13p. *Nucleic Acids Res* **28**, 4733-41.
- Wang, R.C., A. Smogorzewska, et T. de Lange. (2004). Homologous recombination generates T-loop-sized deletions at human telomeres. *Cell* **119**, 355-68.
- Wang, S.S., et V.A. Zakian. (1990). Sequencing of *Saccharomyces* telomeres cloned using T4 DNA polymerase reveals two domains. *Mol Cell Biol* **10**, 4415-9.
- Wang, W., R. Skopp, M. Scofield, et C. Price. (1992). *Euplotes crassus* has genes encoding telomere-binding proteins and telomere-binding protein homologs. *Nucleic Acids Res* **20**, 6621-9.
- Watson, J.D. (1972). Origin of concatemeric T7 DNA. *Nat New Biol* **239**, 197-201.
- Wei, C. et C.M. Price. (2004). Cell cycle localization, dimerization, and binding domain architecture of the telomere protein cPot1. *Mol Cell Biol* **24**, 2091-102.

Wellinger, R.J., K. Ethier, P. Labrecque, et V.A. Zakian. (1996). Evidence for a new step in telomere maintenance. *Cell* **85**, 423-33.

Wellinger, R.J. et D. Sen. (1997). The DNA structures at the ends of eukaryotic chromosomes. *Eur J Cancer* **33**, 735-49.

Wellinger, R.J., A.J. Wolf, et V.A. Zakian. (1993a). Origin activation and formation of single-strand TG1-3 tails occur sequentially in late S phase on a yeast linear plasmid. *Mol Cell Biol* **13**, 4057-65.

-. (1993b). Saccharomyces telomeres acquire single-strand TG1-3 tails late in S phase. *Cell* **72**, 51-60.

Wellinger, R.J. et V.A. Zakian. (1989). Lack of positional requirements for autonomously replicating sequence elements on artificial yeast chromosomes. *Proc Natl Acad Sci U S A* **86**, 973-7.

Weng, N.P., K.S. Hathcock, et R.J. Hodes. (1998). Regulation of telomere length and telomerase in T and B cells: a mechanism for maintaining replicative potential. *Immunity* **9**, 151-7.

Will, C.L. et R. Lührmann. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol* **13**, 290-301.

Williams, K.R., S. Reddigari, et G.L. Patel. (1985a). Identification of a nucleic acid helix-destabilizing protein from rat liver as lactate dehydrogenase-5. *Proc Natl Acad Sci U S A* **82**, 5260-4.

Williams, K.R., K.L. Stone, M.B. LoPresti, B.M. Merrill, et S.R. Planck. (1985b). Amino acid sequence of the UP1 calf thymus helix-destabilizing protein and its homology to an analogous protein from mouse myeloma. *Proc Natl Acad Sci U S A* **82**, 5666-70.

Wright, W.E. et L. Hayflick. (1975). Nuclear control of cellular aging demonstrated by hybridization of anucleate and whole cultured normal human fibroblasts. *Exp Cell Res* **96**, 113-21.

Wright, W.E. et J.W. Shay. (2002). Historical claims and current interpretations of replicative aging. *Nat Biotechnol* **20**, 682-8.

- Wright, W.E., V.M. Tesmer, K.E. Huffman, S.D. Levene, et J.W. Shay. (1997). Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev* **11**, 2801-9.
- Wurtz, J.M., W. Bourguet, J.P. Renaud, V. Vivat, P. Chambon, D. Moras, et al. (1996). A canonical structure for the ligand-binding domain of nuclear receptors. *Nat Struct Biol* **3**, 87-94.
- Xiao, S.H. et J.L. Manley. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* **11**, 334-44.
- Yang, X., M.R. Bani, S.J. Lu, S. Rowan, Y. Ben-David, et B. Chabot. (1994). The A1 and A1B proteins of heterogeneous nuclear ribonucleoparticles modulate 5' splice site selection in vivo. *Proc Natl Acad Sci U S A* **91**, 6924-8.
- Yan-Sanders, Y., G.J. Hammons, et B.D. Lyn-Cook. (2002). Increased expression of heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP) in pancreatic tissue from smokers and pancreatic tumor cells. *Cancer Lett* **183**, 215-20.

Ye, J.Z., J.R. Donigian, M. Van Overbeek, D. Loayza, Y. Luo, A.N. Krutchinsky, et al.

(2004a). TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J Biol Chem.* **279**, 47264-71.

Ye, J.Z., D. Hockemeyer, A.N. Krutchinsky, D. Loayza, S.M. Hooper, B.T. Chait, et al.

(2004b). POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes Dev* **18**, 1649-54.

Yeager, T.R., A.A. Neumann, A. Englezou, L.I. Huschtscha, J.R. Noble, et R.R. Reddel.

(1999). Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukemia (PML) body. *Cancer Res* **59**, 4175-9.

Yoshimura, S.H., H. Maruyama, F. Ishikawa, R. Ohki, et K. Takeyasu. (2004). Molecular

mechanisms of DNA end-loop formation by TRF2. *Genes Cells* **9**, 205-18.

Zahler, A.M., J.R. Williamson, T.R. Cech, et D.M. Prescott. (1991). Inhibition of

telomerase by G-quartet DNA structures. *Nature* **350**, 718-20.

Zakian, V.A. (1996). Structure, function, and replication of *Saccharomyces cerevisiae*

telomeres. *Annu Rev Genet* **30**, 141-72.

- Zamore, P.D. et M.R. Green. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A* **86**, 9243-7.
- Zamore, P.D., J.G. Patton, et M.R. Green. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-14.
- Zhang, L., W. Zhou, V.E. Velculescu, S.E. Kern, R.H. Hruban, S.R. Hamilton, et al. (1997). Gene expression profiles in normal and cancer cells. *Science* **276**, 1268-72.
- Zhang, M., P.D. Zamore, M. Carmo-Fonseca, A.I. Lamond, et M.R. Green. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc Natl Acad Sci U S A* **89**, 8769-73.
- Zhou, J., K. Hidaka, et B. Futcher. (2000). The Est1 subunit of yeast telomerase binds the Tlc1 telomerase RNA. *Mol Cell Biol* **20**, 1947-55.
- Zhou, J., J.L. Mulshine, E.J. Unsworth, F.M. Scott, I.M. Avis, M.D. Vos, et al. (1996). Purification and characterization of a protein that permits early detection of lung

cancer. Identification of heterogeneous nuclear ribonucleoprotein-A2/B1 as the antigen for monoclonal antibody 703D4. *J Biol Chem* **271**, 10760-6.

Zhou, Z., L.J. Licklider, S.P. Gygi, et R. Reed. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182-5.

Zhu, X.D., B. Kuster, M. Mann, J.H. Petrini, et T. de Lange. (2000). Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. *Nat Genet* **25**, 347-52.

Zou, Y., A. Sfeir, S.M. Gryaznov, J.W. Shay, et W.E. Wright. (2004). Does a sentinel or a subset of short telomeres determine replicative senescence? *Mol Biol Cell* **15**, 3709-18.

ANNEXE

Préambule

En guise d'annexe, voici une liste des diverses publications, sous forme de revue, de guide technique ou de chapitre de livre, dans lesquelles j'ai participé à l'élaboration des contenus et à la rédaction. Deux de ces publications ont été jointes, en guise d'exemples.

- A.T. Dandjinou, I. Dionne, S. Gravel, **C. LeBel**, J. Parenteau, et R.J. Wellinger (1999) *Histol Histopathol* **14** : 517-524.
- B. Chabot, **C. LeBel**, S. Hutchison, F.H. Nasim, et M.J. Simard (2003) *Prog. Mol. Subcell. Biol.* Vol. 31 Philippe Jeanteur (Ed.), © Springer-Verlag Berlin Heidelberg : 59-88.
- **C. LeBel** et R.J. Wellinger (2004) *Medecine/Sciences* **20**: 207-212.
- **C. LeBel**, M. Larrivée, A. Bah, N. Laterreur, N. Lévesque, et R.J. Wellinger (2004) *Methods Mol. Biol.* Humana Press, sous presse.
- **C. LeBel** et R.J. Wellinger (2005) *J. Cell Sci.*, **118** (Pt13), 2785-8.

Les versions publiées des articles mentionnés dans la thèse sont insérés en annexe.

Stephen Hutchison, Catherine Le Bel, Marco Blanchette, et Benoit Chabot. (2002) Distinct Sets of Adjacent Heterogeneous Nuclear Ribonucleoprotein (hnRNP) A1/A2 Binding Sites Control 5' Splice Site Selection in the hnRNP A1 mRNA Precursor. *The Journal of Biological Chemistry* 277(33): 29745-29752.

Véronique Bériault, Jean-François Clément, Kathy Lévesque, Catherine Le Bel, Xiao Yong, Benoit Chabot, Eric A. Cohen, Alan W. Cochrane, William F. C. Rigby, et Andrew J. Mouland. (2004) A Late Role for the Association of hnRNP A2 with the HIV-1 hnRNP A2 Response Elements in Genomic RNA, Gag, and Vpr Localization. *The Journal of Biological Chemistry* 279 (42): 44141-44153

Michel Larrivée, Catherine Le Bel, et Raymund J. Wellinger. (2004) The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. *Genes Dev.* 18: 1391-1396.

C. LeBel et R.J. Wellinger (2004) *Medecine/Sciences* **20**: 207-212.

C. LeBel et R.J. Wellinger (2005) *J. Cell Sci.*, **118** (Pt13), 2785-8.

STRUCTURE TERMINALE DES CHROMOSOMES : QUE SE PASSE-T-IL SOUS LE CAPUCHON ?

Mini-Synthèse rédigée par:

Catherine LeBel et Raymund J. Wellinger *

Département de Microbiologie et Infectiologie
Faculté de Médecine
Université de Sherbrooke
3001, 12^e Avenue Nord
Sherbrooke (Québec) J1H 5N4
Canada

* Pour informations et tirés à part; courriel : raimund.wellinger@usherbrooke.ca

Les travaux dans notre laboratoire sont soutenus par des octrois des Instituts de recherche en santé du Canada (IRSC) et l'Institut national du cancer du Canada (INCC). C.L. était boursière du Conseil de recherches en sciences naturelles et en génie (CRNSG) et R.J.W. est un chercheur national du Fonds de la recherche en santé du Québec (FRSQ).

STRUCTURE TERMINALE DES CHROMOSOMES :
QUE SE PASSE-T-IL SOUS LE CAPUCHON ?

03-05-30 // 03-06-02

Mini-Synthèse rédigée par:

Catherine LeBel et Raymund J. Wellinger *

Département de Microbiologie et Infectiologie
Faculté de Médecine
Université de Sherbrooke
3001, 12^e Avenue Nord
Sherbrooke (Québec) J1H 5N4
Canada

* Pour informations et tirés à part; courriel : raimund.wellinger@usherbrooke.ca

Les travaux dans notre laboratoire sont soutenus par des octrois des Instituts de recherche en santé du Canada (IRSC) et l'Institut national du cancer du Canada (INCC). C.L. était boursière du Conseil de recherches en sciences naturelles et en génie (CRNSG) et R.J.W. est un chercheur national du Fonds de la recherche en santé du Québec (FRSQ).

RÉSUMÉ :

Les télomères sont les structures particulières présentes aux extrémités des chromosomes eucaryotes. Ils sont comparables à des ferrets protecteurs de lacets; lorsque le ferret s'use et que le lacet se détériore, on en dispose. Même phénomène pour les chromosomes; lorsque leurs extrémités protectrices sont trop dégradées et atteignent une longueur critique, l'intégrité des chromosomes est altérée et la cellule entre en sénescence. Donc, les télomères, et particulièrement leur structure terminale, sont essentiels pour la stabilité du génome.

Les études du généticien H. J. Muller sur les chromosomes de la mouche à fruits (drosophile) ont donné naissance au concept de télomère en 1938. En analysant les chromosomes de mouches irradiées aux rayons-X, il a remarqué que des délétions ou inversions dans les régions terminales des chromosomes étaient pratiquement introuvables, alors que ces événements étaient détectables dans les autres parties du génome [1]. Muller a donc conclu que les extrémités des chromosomes, qu'il a nommées télomères, devaient posséder une structure particulière qui permettait de conférer une stabilité aux chromosomes. Des études complémentaires de B. McClintock sur le maïs ont par ailleurs démontré que des chromosomes brisés étaient capables de fusionner entre eux, contrairement aux extrémités des chromosomes qui elles demeuraient stables [2]. Selon ces études, les télomères posséderaient des fonctions importantes pour assurer la protection de l'intégrité des chromosomes. Plus récemment, il a été proposé que les télomères agissent comme "capuchon" des chromosomes, les protégeant, entre autres, contre des fusions bout-à-bout et des mécanismes de recombinaison par la machinerie de réparation, et contre des phénomènes de dégradation par des exonucléases. Ce nouveau concept de capuchon chromosomique est maintenant établi dans la communauté scientifique, même si le chemin qui y a mené s'est avéré plutôt sinueux.

Il a d'abord été suggéré en 1972 que la machinerie de réplication conventionnelle est incapable de synthétiser l'extrémité de molécules linéaires d'ADN, que sont les chromosomes eucaryotes. Pour cette raison, les télomères des chromosomes linéaires souffriront des pertes de leurs séquences terminales au fil des cycles de réplication. Étant donné que les télomères sont des structures essentielles pour la protection et la stabilité des chromosomes, leur élimination graduelle suite à chaque ronde de réplication pourrait causer une perte des séquences et des protéines constituant le capuchon protecteur, et ainsi entraîner une instabilité du génome. En effet, chez les humains, il a été observé que les chromosomes raccourcissent dans la plupart des cellules somatiques au fil des années, et il est donc proposé que les télomères jouent un rôle déterminant dans l'apparition de la sénescence cellulaire. Cependant, il existe une enzyme spécifique, la télomérase, qui a d'abord été identifiée chez le cilié *Tetrahymena*, qui peut prévenir ce raccourcissement naturel des chromosomes, en ajoutant des répétitions télomériques à l'extrémité des chromosomes [3]. Chez les humains, cette enzyme se retrouve activée dans les cellules germinales, les cellules transformées et dans 90% des cellules cancéreuses [4]. Ces dernières découvertes ont créé un réel engouement car elles suggèrent un lien entre l'activation de la télomérase et la carcinogénèse. Vu que cette enzyme n'est pas présente dans la plupart des cellules humaines somatiques normales et qu'elle semble activée spécifiquement dans la majorité des cellules cancéreuses, des thérapies anti-cancer basées sur l'inhibition de la télomérase pourraient donc être applicables à une panoplie de cancers.

De nombreux efforts de recherche investis principalement sur la télomérase ont jeté une ombre sur les recherches visant à mieux comprendre les caractéristiques du capuchon télomérique, dont les mécanismes moléculaires de protection des chromosomes sont encore mal définis. Grâce au séquençage de l'ADN télomérique chez certaines espèces,

la structure primaire de l'ADN présent aux extrémités des chromosomes est connue et se révèle relativement bien conservée entre les espèces (figure 1). Pour la plupart, l'ADN télomérique est constitué de courtes séquences, répétées en tandem [5]. Le nombre de répétitions varie entre les espèces, entre les individus et même entre les chromosomes d'une même cellule. Fait important à noter chez tous les organismes étudiés, le brin riche en guanines (G-riche), qui s'oriente de 5' à 3' vers la fin du chromosome, dépasse le brin complémentaire riche en cytosines (C-riche), créant ainsi une extension simple-brin en 3' du chromosome (figure 1). Ironiquement, la drosophile est la seule espèce qui fait exception aux règles générales de constitution des télomères, possédant plutôt, aux extrémités des chromosomes, un nombre multiple de copies de rétrotransposons télomériques spécifiques. On pense que les télomères de drosophile dérivent des mêmes télomères ancestraux que ceux qui ont donné naissance à des télomères réguliers chez d'autres espèces cousines de la drosophile. De ce fait, les télomères de drosophile seraient des modifications des séquences usuelles, plutôt que des réinventions radicales de la structure télomérique.

La structure terminale de l'ADN possédant une extension télomérique G-riche, semble donc bien conservée à travers le règne eucaryote, lui prédisant ainsi une importance fonctionnelle. Il est suggéré en effet que l'extension est essentielle pour assurer la fonction de capuchon télomérique [6, 7]. Il est aussi probable que cette structure terminale bien définie de l'ADN puisse en fait être responsable des fonctions de protection de l'intégrité du génôme, énoncées lors de la première définition du capuchon télomérique, il y a 60 ans. Le texte qui suit met donc l'accent sur les récents développements concernant l'analyse et l'importance de la structure terminale de l'ADN chromosomique chez différents organismes.

Les débuts chez les ciliés

La première séquence télomérique à être déterminée fut celle du protozoaire cilié *Tetrahymena thermophila* [8]. Les télomères de cet organisme comportent 50 à 70 répétitions de la séquence TTGGGG/CCCCAA (T_2G_4/C_4A_2), et le brin G-riche s'oriente de 5' à 3' vers l'extrémité du chromosome. Au début des années 1980, la taille de l'extension 3' simple-brin a été estimée entre 12 et 16 nucléotides (figure 1). Cette extension est détectée durant tout le cycle cellulaire, et est présente sur chacun des télomères de l'abondant ADN macronucléaire [9]. Il a été par la suite démontré que la taille exacte des extensions, ainsi que leurs séquences, sont régulées très strictement chez les ciliés. La majorité des extensions se terminent par la séquence 5'-TTGGGGT-3' et plus de 80% de ces extensions sont de 14-15 ou 20-21 nucléotides [10]. De plus, il a été confirmé que ces extensions sont présentes durant tout le cycle cellulaire, mais qu'elles peuvent varier lorsque les cellules sont soumises à certaines restrictions nutritionnelles ou lorsqu'elles sortent du cycle cellulaire. Finalement, certaines expériences suggèrent fortement que, suite à la réplication de l'ADN, des extensions G-riche peuvent être présentes sur tous les télomères, aux deux extrémités des chromosomes (figure 2) [10].

Les levures : petites variations sur le thème

Chez la levure *Saccharomyces cerevisiae*, il existe une certaine dynamique dans la régulation de la taille des extensions selon les phases du cycle cellulaire. En accord avec

les résultats obtenus chez les ciliés, nos résultats récents suggèrent que les télomères de levures possèdent aussi une extension de 15-20 bases pendant la majorité du cycle cellulaire (données non-publiées). Par contre, les chromosomes de *S. cerevisiae* acquièrent des extensions G-riches de plus de 30 nucléotides, de façon transitoire à la fin de la phase S [11]. D'autres évidences suggèrent fortement que ces longues extensions peuvent aussi être présentes sur tous les télomères, suite à la réplication de l'ADN, donc à toutes les extrémités des chromosomes [12]. Par contre, au contraire de ce qui est observé chez les ciliés, la séquence 3' terminale de la levure peut être différente pour chaque télomère; bien qu'elle corresponde toujours à la séquence consensus TG₁₋₃, la portion 3' terminale peut être hétérogène, due à la variation du nombre de G pouvant être présent dans chacune des répétitions[13]. La dynamique des extensions simple-brin à travers le cycle cellulaire et la présence d'une séquence spécifique à la toute fin des chromosomes semblent donc être des événements moléculaires qui sont variables selon les espèces. Néanmoins, l'ensemble des résultats obtenus chez les ciliés et les levures suggèrent que la protection des extrémités des chromosomes est assurée par une structure terminale définie, comprenant une courte extension simple brin (14-20 nucléotides) et des protéines spécifiques associées, ce qui forme le capuchon télomérique.

Les cellules mammifères : vers un changement du mode de pensée ?

Pour les cellules de plantes et de mammifères, les études ont montré que les séquences télomériques et les extensions simple-brin s'avèrent plus longues que pour les organismes décrits jusqu'à présent. En effet, les télomères des cellules humaines possèdent une extension G-riche d'environ 150-200 bases, qui serait aussi présente durant toutes les phases du cycle cellulaire [14-16]. Une controverse scientifique survient cependant quant à la distribution de ces extensions aux deux extrémités d'un même chromosome. En utilisant une technique appelée PENT (*primer extension/nick translation*), Makarov *et al.* ont montré la présence d'extensions G-riches sur plus de 80% des extrémités des chromosomes humains [14]. Ce résultat est en désaccord avec le modèle de réplication conventionnelle (figure 2), qui prédit que seulement 50% des extrémités (produites par la réplication du brin retardé) peuvent posséder une extension 3' simple-brin, suite à l'enlèvement de la dernière amorce d'ARN ayant servi à la synthèse du fragment d'Okazaki (hybride ARN-ADN synthétisé par la polymérase α /primase pour permettre la réplication par le brin retardé). Donc, les extensions qui sont détectées sur les extrémités répliquées par le brin avancé pourraient être dues à des activités de dégradation aux télomères, activités qui ont aussi été mises en évidence dans la levure *S. cerevisiae* [12]. Par contre, comme le moment d'action précis de la télomérase n'est pas encore bien défini, il se peut aussi que ces extensions soient créées par le rallongement du brin avancé par la télomérase, qui agirait suite à la réplication conventionnelle, en utilisant comme substrat une très courte extension laissée par la machinerie de réplication ou générée par des exonucléases.

Wright *et al.* ont obtenu chez l'humain des résultats qui s'avèrent plus comparables à ceux obtenus chez les plantes, et qui sont en accord avec le modèle de réplication conventionnelle. Selon les auteurs, les chromosomes humains de fibroblastes peuvent être asymétriques, avec une longue extension d'un côté et une très courte extension de

l'autre [16, 17]. Par microscopie électronique, ils ont évalué la taille des longues extensions simple-brin à environ 200 ± 75 nucléotides, mais la taille des courtes extensions demeure inconnue [16]. Donc, le simple fait que les extensions sont requises pour le bon fonctionnement de la protection au niveau des télomères indique que, peu importe la taille des extensions, celles-ci devraient être présentes sur tous les télomères [7].

Ces résultats obtenus chez l'humain sont en accord avec un modèle "linéaire" des extrémités des chromosomes. Ce modèle stipule que l'extension d'ADN simple-brin et les protéines spécifiques associées aux télomères forment le capuchon télomérique. Cependant, en 1999, un changement radical de ce mode de pensée classique concernant la protection des chromosomes est survenu. Il a été montré pour la première fois par microscopie électronique que les extrémités des chromosomes de mammifères seraient en fait repliées sur elles-mêmes pour former une énorme boucle terminale, appelée t-loop [18]. Ces boucles sont créées grâce à l'invasion de l'extrémité 3' simple-brin à l'intérieur de la portion double-brin du télomère (figure 3). De cette façon, il y a déplacement d'une partie de l'ADN télomérique double-brin, qui devient simple-brin à l'intérieur de la boucle même, formant une D-loop [18]. Grâce à des extensions terminales simple-brin relativement longues chez les mammifères, et aussi à l'implication potentielle de certaines protéines télomériques spécifiques, ces structures terminales permettraient aux cellules de protéger leurs extrémités chromosomiques, et ainsi de les différencier des cassures double-brin qui surviennent ailleurs dans l'ADN. Étant donné le contenu riche en G des extensions télomériques, des structures alternatives pourraient aussi se former aux extrémités, comme par exemple des duplexes G-G, de même que des structures entre 4 brins d'ADN telles les quadruplexes de G (G-quartets ou G4-DNA) [5].

De plus, de récents résultats publiés par le laboratoire de Weinberg ont apporté des évidences additionnelles concernant l'importance des extensions télomériques simple-brin. En utilisant une technique de ligation d'oligos télomériques complémentaires à l'extension G-riche (T-OLA), mise au point par le laboratoire d'Ambrosio [19], ils ont réussi à montrer une corrélation entre l'érosion des extensions 3' simple-brin et l'entrée en sénescence des cellules humaines. Cette érosion semble être le résultat direct des divisions cellulaires continues, plutôt qu'être une conséquence de la sénescence. Ces évidences suggèrent qu'une altération spécifique dans la structure télomérique (et plus spécifiquement de l'extension G-riche) survient lors de la sénescence et que c'est ce changement précis, plutôt que la taille globale des télomères, qui déclenche la sénescence cellulaire [20].

De retour chez les ciliés : deux modes de protection ?

Suite à la découverte des structures de boucles chez les mammifères, des structures t-loop semblables ont pu être observées sur l'ADN micronucléaire d'*Oxytricha fallax*, qui possède de longues extensions pouvant permettre la formation de ces boucles terminales. Par contre, l'ADN télomérique des molécules macronucléaires du même organisme possède des extensions trop courtes pour former des t-loops, et il se retrouve donc sous forme linéaire simple-brin, protégé par des complexes spécifiques d'hétérodimères protéiques [21]. Donc, chez cet organisme, selon la provenance de

l'ADN qui est étudié, deux mécanismes différents de protection des chromosomes pourraient être utilisés en même temps et ce, par une même cellule. Des structures de t-loops ont aussi été observées chez *Trypanosoma brucei*, aux deux extrémités des chromosomes [22]. Cette dernière observation rejoint celle qui a été faite par Makarov *et al.* chez les mammifères, concernant la présence d'extensions simple-brin sur les deux brins d'ADN nouvellement synthétisés. Donc, la présence de longues extensions simple-brin semble importante pour permettre la création des t-loops dans tous les organismes.

Conclusion :

L'ensemble de ces résultats suggère qu'il y aurait en fait deux arrangements possibles aux extrémités des chromosomes eucaryotes : les t-loops qui requièrent de longues extensions afin de permettre la formation même de la boucle, et les courtes extensions simple-brin qui sont protégées par des complexes protéiques spécifiques. Ces deux arrangements sont-ils nécessairement exclusifs ? Il semble que non. Récemment, une nouvelle protéine télomérique protectrice a été découverte chez la levure de fission *S. pombe* et chez les humains. Cette protéine possède des caractéristiques très similaires à celles des protéines spécifiques qui protègent les courtes extensions chez les ciliés et chez la levure de bourgeonnement *S. cerevisiae* [23]. Comme les extensions impliquées dans la formation de la structure de t-loop sont relativement longues, il serait possible qu'une courte séquence à la toute fin de ces extensions demeure sous forme simple-brin dans la boucle, non associée au brin complémentaire (figure 3). Cette courte extension simple-brin pourrait lier des protéines protectrices conservées chez les eucaryotes, et pourrait être la condition *sine qua non* pour la formation du capuchon télomérique. Il est donc possible que la fonction de protection des chromosomes eucaryotes associée au capuchon terminal requière un complexe nucléoprotéique conservé entre les espèces. La signification biologique et le mécanisme de formation des t-loops sont sources de questions brûlantes en ce moment, mais il se peut encore que d'autres surprises puissent être cachées sous le capuchon...

Références :

1. Muller HJ. The remaking of chromosomes. *The Collecting Net* 1938 ; 13 : 181-195, 198.
2. McClintock B. The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proc Natl Acad Sci USA* 1939 ; 25 : 405-416.
3. Greider CW, Blackburn EH. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* 1985 ; 43 : 405-413.
4. Kim NW, Piatyszek MA, Prowse KR *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* 1994 ; 266 : 2011-2015.
5. Wellinger RJ et Sen D. The DNA structures at the ends of eukaryotic chromosomes. *Eur J Cancer* 1997 ; 33 : 735-749.

6. Sandell LL, Zakian VA. Loss of a yeast telomere: arrest, recovery, and chromosome loss. *Cell* 1993 ; 75 : 729-739.
7. van Steensel B, Smogorzewska A, de Lange T. TRF2 protects human telomeres from end-to-end fusions. *Cell* 1998 ; 92 : 401-413.
8. Blackburn EH, Gall JG. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J Mol Biol* 1978 ; 120 : 33-53.
9. Klobutcher LA, Swanton MT, Donini P, Prescott DM. All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc Natl Acad Sci USA*. 1981 ; 78 : 3015-3019.
10. Jacob NK, Skopp R, Price CM. G-overhang dynamics at *Tetrahymena* telomeres. *Embo J* 2001 ; 20 : 4299-4308.
11. Wellinger RJ, Wolf AJ, Zakian VA. Saccharomyces telomeres acquire single-strand TG1-3 tails late in S phase. *Cell* 1993 ; 72 : 51-60.
12. Wellinger RJ, Éthier K, Labrecque P, Zakian VA. Evidence for a new step in telomere maintenance. *Cell* 1996 ; 85 : 423-433.
13. Forstemann K, Hoss M, Lingner J. Telomerase-dependent repeat divergence at the 3' ends of yeast telomeres. *Nucleic Acids Res* 2000 ; 28 : 2690-2694.
14. Makarov VL, Hirose Y, Langmore JP. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* 1997 ; 88 : 657-666.
15. McElligott R, Wellinger RJ. The terminal DNA structure of mammalian chromosomes. *Embo J* 1997 ; 16 : 3705-3714.
16. Wright WE, Tesmer VM, Huffman KE, Levene SD, Shay JW. Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev* 1997 ; 11 : 2801-2809.
17. Riha K, McKnight TD, Fajkus J, Vyskot B, Shippen DE. Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. *Plant J* 2000 ; 23 : 633-641.
18. Griffith JD, Comeau L, Rosenfield S, *et al.* Mammalian telomeres end in a large duplex loop. *Cell* 1999 ; 97 : 503-514.
19. Cimino-Reale G, Pascale E, Battiloro E, *et al.* The length of telomeric G-rich strand 3'-overhang measured by oligonucleotide ligation assay. *Nucleic Acids Res* 2001 ; 29 : e35.
20. Stewart SA, Ben-Porath I, Carey VJ, *et al.* Erosion of the telomeric single-strand overhang at replicative senescence. *Nat. Genet.* 2003 ; 33 : 492-496.
21. Murti KG, Prescott DM. Telomeres of polytene chromosomes in a ciliated protozoan terminate in duplex DNA loops. *Proc Natl Acad Sci USA* 1999 ; 96 : 14436-14439.
22. Munoz-Jordan JL, Cross GA, deLange T, Griffith JD. t-loops at trypanosome telomeres. *Embo J* 2001 ; 20 : 579-588.
23. Baumann P, Cech TR. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* 2001 ; 292 : 1171-1175.

Figure 1 A : Visualisation de la structure primaire de l'ADN d'un télomère de l'organisme modèle *Tetrahymena thermophila*

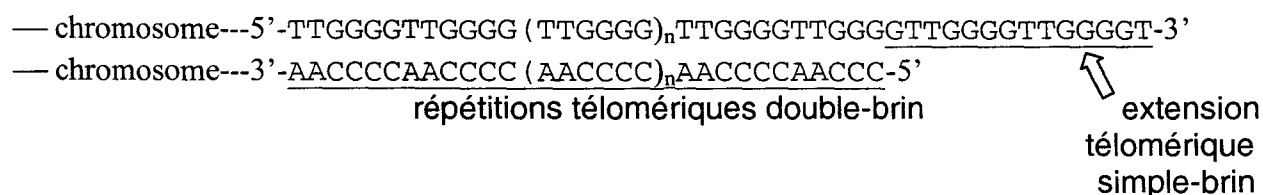


Figure 1 B : Séquences télomériques du brin 5' à 3' vers l'extrémité du chromosome, dans certains organismes sélectionnés [5, 17].

Organismes	Séquences répétées / abréviation des répétitions
Protozoaires	
<i>Tetrahymena</i>	TTGGGG / T ₂ G ₄
<i>Oxytricha</i>	TTTTGGGG / T ₄ G ₄
<i>Euplotes</i>	TTTTGGGG / T ₄ G ₄
<i>Trypanosoma</i>	TTAGGG / T ₂ AG ₃
Champignons	
<i>Saccharomyces</i>	TGTGGTGTGGG / (TG) ₁₋₆ TG ₂₋₃
Plantes	
<i>Arabidopsis</i>	TTTAGGG / T ₃ AG ₃
<i>Silene</i>	TTTAGGG / T ₃ AG ₃
Vertébrés	
<i>Homo Sapiens</i>	TTAGGG / T ₂ AG ₃
Autres vertébrés	TTAGGG / T ₂ AG ₃

Légende des figures :

Figure 2.

Le problème du brin avancé lors de la réplication des télomères. La réplication semi-conservative de l'ADN ne permet pas de régénérer une extension 3' simple-brin du côté du brin avancé. Par contre, du côté du brin retardé l'enlèvement de la dernière amorce d'ARN (rouge), laisse une courte extension. Donc, de ce côté-ci, un allongement des répétitions télomériques par la télomérase est possible avec ou sans modifications additionnelles par des exonucléases.

L'ADN parental est montré en noir, et l'ADN nouvellement synthétisé en bleu. Les flèches indiquent la polarité de l'ADN. Il est encore incertain si la télomérase ajoute les répétitions télomériques avant ou après la réplication conventionnelle.

The leading strand problem of telomere replication. Semi-conservative DNA replication does not allow the regeneration of the 3' overhang on leading-strand ends. On lagging-strand ends, the removal of the last RNA primer (in red) generates a short extension which could serve to recruit the telomerase enzyme. Therefore an extension of the telomeric repeats on such ends could occur with or without processing by exonucleases. Parental DNA is shown in black, and the newly replicated DNA is shown in blue. Arrows indicate the polarity of DNA. It is still unknown if telomerase adds telomeric repeats before or after the conventional replication.

Figure 3.

A) Vue classique d'une extrémité de chromosome avec une courte extension en 3' et un complexe protéique protégeant l'ADN chromosomal (modèle linéaire).

B) Les télomères des mammifères se terminent par de larges boucles, les t-loops. Le double-brin télomérique se replie sur lui-même et l'extension 3' simple-brin s'insère dans la portion double-brin, ce qui génère la D-loop (displacement loop). La protéine TRF2 se retrouve liée à la base du D-loop in vitro. Les t-loops, en association avec un complexe protéique terminal, pourraient permettre de distinguer les extrémités des chromosomes des cassures double-brin dans l'ADN, et ainsi éviter la détection par les points de contrôle ou mécanismes de contrôle du cycle cellulaire, et la machinerie de réparation des cassures d'ADN.

Les régions simple brin du t-loop et les extensions en 3' sont indiquées en rouge.

A) Standard view of the DNA arrangement at chromosome ends. The short 3' overhang is bound by a protective protein complex (linear model).

B) Mammalian telomeres loop back on themselves to form a lariat structure, where the 3' single-strand extension invades the duplex portion of telomeric repeats to form the D-loop (displacement loop). TRF2 is found at the junction of the lariat in vitro and helps to create and/or stabilize the t-loop. Together with a terminal protein complex, t-loops could allow protection of the chromosome ends from detection by cell cycle checkpoints and DNA break repair activities.

Single-stranded regions of the t-loop and the 3' extensions are indicated in red.

Figure 2

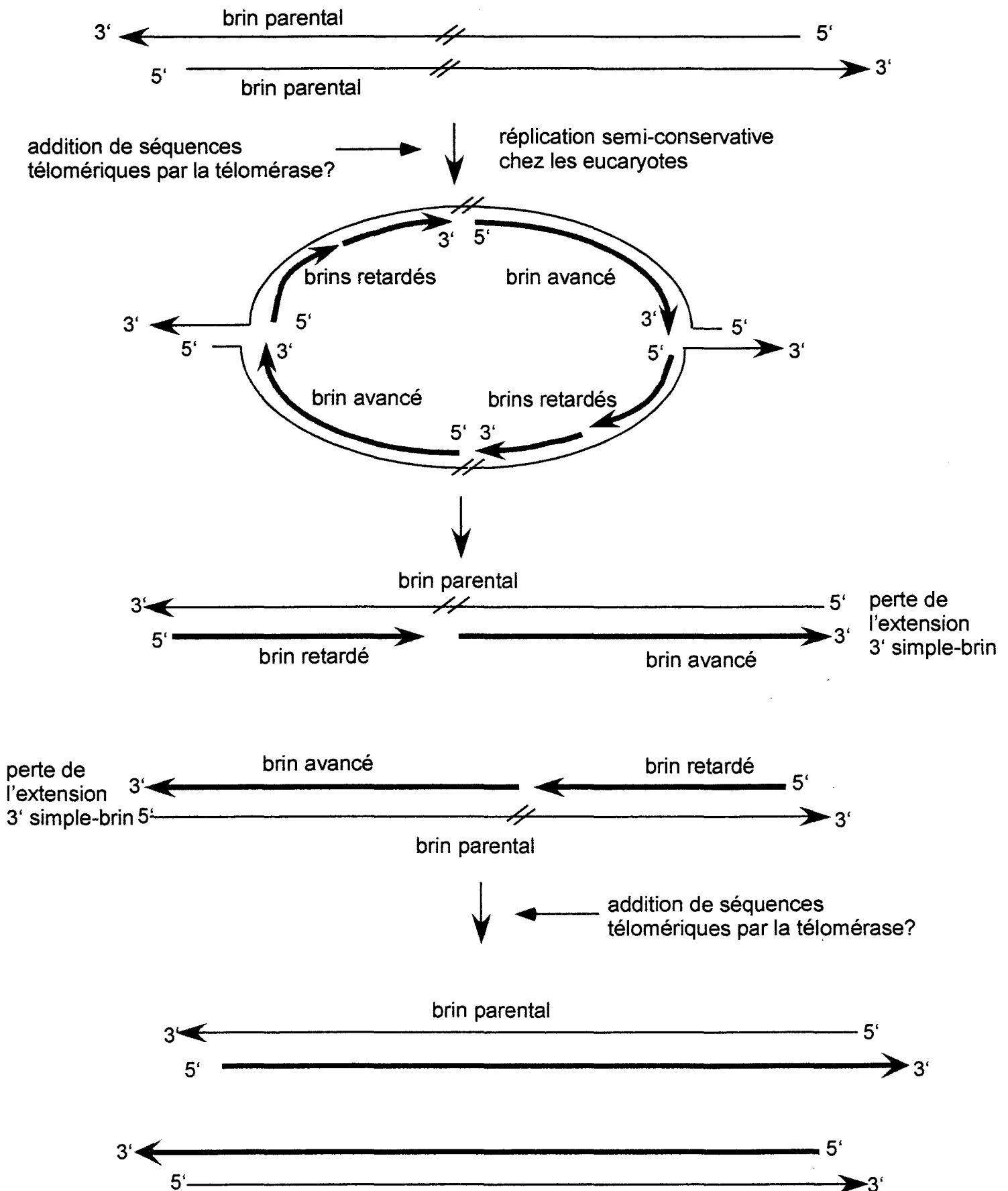
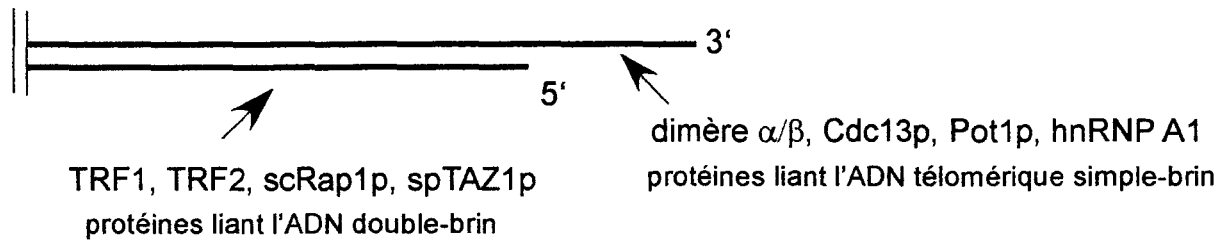
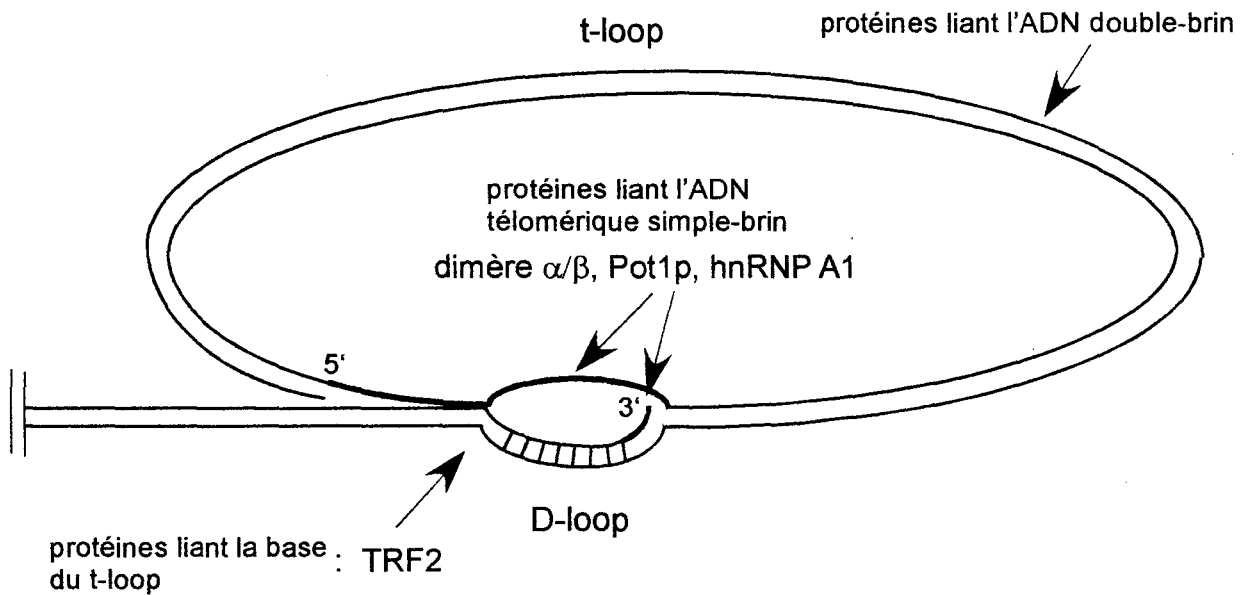


Figure 3.

A) Le modèle linéaire classique



B) Le modèle révisé de t-loop



Cell Science at a Glance:

Telomeres: what's new at your end?

Catherine LeBel and Raymund J. Wellinger*
Dept. of Microbiology and Infectious Diseases
Université de Sherbrooke
Sherbrooke, QC, J1H 5N4
Canada

* Author for Correspondence (Email: Raymund.Wellinger@Usherbrooke.ca)

The ends of linear eukaryotic chromosomes, the telomeres, fulfill unique and essential functions for genome integrity. First and above all, there is the fact that chromosome ends are also ends of the underlying DNA molecules. In general, non-telomeric double-strand DNA breaks (DSB) are not tolerated and are rapidly repaired. This scenario however does not apply to telomeres (Muller, 1938; McClintock, 1939). Indeed, telomeric chromosome ends are bound by an increasingly complex and interesting cast of factors that ensure this distinction, hence protecting chromosome ends from degradation or fusion (d'Adda di Fagagna et al., 2004). Secondly, due to the semi-conservative synthesis of DNA, a DNA-end can not be completely duplicated (Watson, 1972; Olovnikov, 1973). In order to avoid continuous sequence losses from the telomeres in dividing cells, special mechanisms have evolved. In most eukaryotic organisms, the solution to the problem involves a ribonuclear complex, called telomerase, that is minimally composed of a reverse transcriptase catalytic protein subunit and an RNA subunit that is used as template (Greider and Blackburn, 1987). Recent research into how these two functions are accommodated at telomeres and how they are integrated into the cell division cycle have revealed that the overall telomere structure is dynamic with dramatic changes occurring particularly in S-phase (Blackburn, 2001; Vega et al., 2003; Smogorzewska and de Lange, 2004). Telomere structure and function are most intensely studied in model systems such as budding yeasts and in human cell lines. Therefore, we chose those organisms to illustrate similarities and differences in telomere composition and structure.

Telomeric DNA structures are highly conserved amongst eukaryotes. In virtually all organisms, chromosomes end with an array of short direct repeats. For example, human telomeres are comprised of 2-20 kb of (TTAGGG)_n repeats. It is quite evident that the species-specific repeats are related to each other, being composed of a G-rich and a C-rich strand (see Table). Moreover, the directionality of the repeats is conserved such that

the 3'-end of the chromosome is on the G-rich strand. Also quite universal is the fact that the G-rich strand is longer than the C-rich strand and forms a single-strand extension at the very ends of the chromosomes (also called G-tails). For species with relatively long constitutive G-tails, such as humans, there is evidence that the terminal G-tails invade and displace G-strands in the double-strand portion of the telomere, yielding a configuration that has been dubbed t-loop (Griffith et al., 1999). Budding yeast telomeres, composed of irregular (TG)₁₋₆TG₂₋₃-repeats, are about 300 bp long and their G-tails normally are short (Larrivee et al., 2004). However, this terminal DNA arrangement is dynamic and changes during the cell cycle: longer G-tails can be detected on yeast telomeres during S-phase, when telomeres are replicated (Wellinger et al., 1993).

With few exceptions, the telomeric repeat tract lengths are not precise but rather centered around a genetically determined mean. Thus, in dividing cell populations, the actual tract length at individual telomeres can fluctuate around the average mean size. Importantly, an adequate, but not precisely determined minimal tract of telomeric repeats is essential for functional telomeres in all species. Finally, in certain plant and insect species, much longer (≥ 1 kb) and complex repeats are found at chromosome ends. Although not discussed in detail here, the mechanism of telomeric DNA-maintenance in these latter organisms clearly are related to the ones active in organisms with the shorter repeats (Pardue and DeBaryshe, 2003).

Telomere-associated proteins identified in various experimental systems clearly serve similar functions. For example, while not being orthologues of each other, the primary double-strand repeat binding proteins TRF1 and TRF2 of humans and the major telomere binding protein of budding yeast, Rap1p, share a similar DNA binding motif and serve as binding platform for a number of additional telomere associated proteins. Moreover, the human orthologue of this yeast Rap1p protein, hRAP1, associates with telomeres via an interaction with TRF2. The major G-tail binding proteins (Cdc13p in budding yeast, POT1 in mammals) also contain related protein motifs and are viewed as functional orthologues.

The first major function of the telomeric complex is to protect the chromosome ends from degradation. Furthermore, telomeres also prevent chromosomal DNA-ends from being recognized as DNA damage. In yeast, Cdc13p is essential for these functions since after an experimentally induced inactivation of it, telomeric DNA is degraded (Garvik et al., 1995). Although an initial report suggests a similar role for POT1 in humans (Veldman et al., 2004), the phenotypes induced by a complete absence of hPOT are not known yet. Therefore, proteins associating with the G-tails appear essential for preventing enzymes to either degrade telomeric DNA or process the DNA-ends such that they would be suitable for double-strand break repair. Proteins associated with the double-strand portion of the repeats are also important for this end-protection function. In particular, the human TRF2 protein is crucial for maintaining a normal chromosomal end-structure as it collaborates in the processes leading to t-loops (van Steensel et al., 1998). However, it is less clear how telomeric proteins physically protect the DNA from DNA-processing events, such as brought about by stochastic nuclease attack. For certain ciliates, the only system where this has been addressed directly, the binding of Pot1-related proteins to telomeres can protect the DNA from degradation, at least *in vitro* (Gottschling and Zakian, 1986). However, we know even less about how the proteins

prevent telomeres from being detected as DNA-damage. This issue is further compounded by the paradoxical finding that proteins that are involved in sensing and processing damaged DNA are also intimately associated with telomeric chromatin. For example, orthologues of the Ku-complex, of ATM/ATR proteins as well as of the Mre11-complex clearly localize to telomeres and/or have specific functions in maintaining a functional telomere in yeasts and humans (d'Adda di Fagagna et al., 2004). However, there is increasing evidence that these proteins may act on telomeres primarily during S-phase, when telomeric DNA is replicated (see below). Thus, their functions on telomeres outside S currently are poorly understood.

Outside of S phase, the telomere complex adopts a fold-back structure that is thought to be organized in a specialized chromatin structure through associations between telomeric and subtelomeric binding proteins. The result is a tight chromatin domain which extends into nucleosomal DNA and represses transcriptional activity in telomere-proximal areas, an effect that has been dubbed telomere position effect (TPE) (Gottschling et al., 1990; Baur et al., 2001). Related to this issue, telomeric DNA in mammals in fact is not exclusively bound by telomere binding proteins; a significant portion of telomere distal repeat DNA is organized in nucleosomes.

The second major problem arising at telomeres is how to maintain telomeric repeat tracts at functional lengths in dividing cell populations. The most prevalent solution for the shortcomings of the conventional replication machinery involves the telomerase holoenzyme, a reverse transcriptase that uses the free 3'-end on the G-tails as primer to add species-specific telomeric G-strand repeats by using a short region on its associated RNA as template (Greider and Blackburn, 1987; Greider and Blackburn, 1989; Kelleher et al., 2002; Smogorzewska and de Lange, 2004). The C-rich telomeric strand is synthesized by the components of the conventional replication machinery that is involved in lagging-strand synthesis. Furthermore, telomeric G-strand and C-strand syntheses may be coordinated and temporally co-regulated (Price, 1997; Diede and Gottschling, 1999; Adams Martin et al., 2000). Consistent with this notion, synthesis of telomeric repeats occurs specifically in late S-phase (Marcand et al., 2000). Therefore, an emphasis of recent research has been on how this telomere-specific end-replication becomes actively engaged and how it is regulated to maintain telomeres at lengths that are within the species-specific window.

Progress towards those ends has been made in several areas. First, with few exceptions, the evidence argues against replication initiating at telomeres. Therefore, replication forks will move from the interior of the chromosome towards the ends. Given the length and conserved directionality of the double-strand telomeric repeats, the C-strand will always be synthesized by lagging-strand synthesis and the G-strand by leading-strand synthesis. Although it is not known exactly how far the conventional fork remains intact on telomeres, the outcome at the very ends on the two daughter chromatids is thought to be different: a blunt end on the leading-strand-ends and a recessed-end at the lagging-strand-ends. Indeed, there is evidence that these two types of ends may either be sensed differently or require different levels of processing (Bailey et al., 2001; Parenteau and Wellinger, 2002). Irrespective of these considerations, it is important to retain the intimate association between the conventional DNA-replication machinery with telomeric repeat replication and hence, telomere length regulation (Chakhparonian and Wellinger, 2003). The passage of the replication fork through the

telomeric repeats may destabilize the telomeric chromatin and remove t-loop structures, allowing for access to telomeres by telomerase and perhaps other enzymes.

Second, the molecular details of the transition from conventional replication to end-replication are just beginning to be understood. Conceivably, telomere-associated DNA-repair proteins could get directly involved at this step. For example, the presumed blunt ends on the leading-strand chromatids could be recognized as damage, engaging the associated signaling and processing proteins such as Ku, the Mre11-complex and ATM. The result would be the generation of the required G-tails and binding of the G-tail associated proteins. In yeast cells, the G-tail binding protein Cdc13p is required to orchestrate and regulate end-replication by telomerase and C-strand synthesis (Evans and Lundblad, 2000; Qi and Zakian, 2000). The link between the DNA-repair proteins and end-replication is further emphasized by the fact that, in yeast, the Ku-proteins directly bind the RNA-component of telomerase and play a crucial role for the association of the enzyme with telomeres (Fisher et al., 2004). At the point of engaging productive end-replication, the regulation of further interactions becomes crucial: the telomerase enzyme must gain access to its substrate, and specific telomerase sub-components, such as Est1p in yeast, must be associated with Cdc13p and telomerase in an active fashion. If these intricate reorganizations at the telomeres do not occur in the restricted temporal window in S-phase, productive end-replication may not be established. In yeast for example, not all telomeres are elongated in every cell cycle and those telomeres with relatively short repeat tracts have a better chance of completing the required steps successfully (Teixeira et al., 2004).

Third, for telomere lengths to stay within certain limits, an active end-elongation complex must be halted, not a small feat. The precise molecular mechanisms involved in this regulatory step are nebulous. However, in yeast and mammals, the telomeric proteins involved in assembling the complexes on the double-strand repeats are crucial for this regulation. Indeed, they somehow progressively inhibit telomeric repeat synthesis as the overall length of the particular telomere increases (Marcand et al., 1997; van Steensel and de Lange, 1997; Shore, 2001; Smogorzewska and de Lange, 2004). In addition to merely stopping the elongation, a functional telomeric complex must be re-generated following synthesis. At the level of DNA, in mammalian cells this may involve the creation of the t-loop structures.

For all of these above steps, one can expect that post-translational modifications of the central proteins will play important, albeit as yet unknown regulatory roles. The observed mean lengths of telomeric repeat tracts thus will be the net result of multiple processes occurring at telomeres, including those affecting conventional replication, end-replication, end-processing steps, accidental degradation and even occasional losses of large portions of the tracts by a recombination based mechanism called telomere rapid deletion (TRD)(Lustig, 2003). All of these mechanisms may be subject to their proper regulatory events, which renders the study of overall telomere length quite complex. However, given the established importance of telomeres for genome stability as well as the strong links between telomere biology, human disease and aging, there really is no end in sight for research on telomeres.

References

- Adams Martin, A., Dionne, I., Wellinger, R. J. and Holm, C.** (2000). The function of DNA polymerase alpha at telomeric G tails is important for telomere homeostasis. *Mol Cell Biol* **20**, 786-96.
- Bailey, S. M., Cornforth, M. N., Kurimasa, A., Chen, D. J. and Goodwin, E. H.** (2001). Strand-specific postreplicative processing of mammalian telomeres. *Science* **293**, 2462-5.
- Baur, J. A., Zou, Y., Shay, J. W. and Wright, W. E.** (2001). Telomere position effect in human cells. *Science* **292**, 2075-7.
- Blackburn, E. H.** (2001). Switching and signaling at the telomere. *Cell* **106**, 661-73.
- Chakhparonian, M. and Wellinger, R. J.** (2003). Telomere maintenance and DNA replication: how closely are these two connected? *Trends Genet* **19**, 439-46.
- d'Adda di Fagagna, F., Teo, S. H. and Jackson, S. P.** (2004). Functional links between telomeres and proteins of the DNA-damage response. *Genes Dev* **18**, 1781-99.
- Diede, S. J. and Gottschling, D. E.** (1999). Telomerase-mediated telomere addition in vivo requires DNA primase and DNA polymerases alpha and delta. *Cell* **99**, 723-33.
- Evans, S. K. and Lundblad, V.** (2000). Positive and negative regulation of telomerase access to the telomere. *J Cell Sci* **113**, 3357-64.
- Fisher, T. S., Taggart, A. K. and Zakian, V. A.** (2004). Cell cycle-dependent regulation of yeast telomerase by Ku. *Nat Struct Mol Biol* **11**, 1198-205.
- Garvik, B., Carson, M. and Hartwell, L.** (1995). Single-stranded DNA arising at telomeres in *cdc13* mutants may constitute a specific signal for the RAD9 checkpoint. *Mol Cell Biol* **15**, 6128-38.
- Gottschling, D. E., Aparicio, O. M., Billington, B. L. and Zakian, V. A.** (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* **63**, 751-62.
- Gottschling, D. E. and Zakian, V. A.** (1986). Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell* **47**, 195-205.
- Greider, C. W. and Blackburn, E. H.** (1987). The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**, 887-98.
- Greider, C. W. and Blackburn, E. H.** (1989). A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* **337**, 331-7.
- Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H. and de Lange, T.** (1999). Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503-14.
- Kelleher, C., Teixeira, M. T., Forstemann, K. and Lingner, J.** (2002). Telomerase: biochemical considerations for enzyme and substrate. *Trends Biochem Sci* **27**, 572-9.
- Larrivee, M., LeBel, C. and Wellinger, R. J.** (2004). The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. *Genes Dev* **18**, 1391-6.
- Lustig, A. J.** (2003). Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. *Nat Rev Genet* **4**, 916-23.
- Marcand, S., Brevet, V., Mann, C. and Gilson, E.** (2000). Cell cycle restriction of telomere elongation. *Curr Biol* **10**, 487-90.
- Marcand, S., Gilson, E. and Shore, D.** (1997). A protein-counting mechanism for telomere length regulation in yeast. *Science* **275**, 986-90.

- McClintock, B.** (1939). The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proc Natl Acad Sci U S A* **25**, 405-416.
- Muller, H. J.** (1938). The remaking of chromosomes. *The Collecting Net* **13**, 181-95, 198.
- Olovnikov, A. M.** (1973). A theory of marginotomy. *J Theor Biol* **41**, 181-190.
- Pardue, M. L. and DeBaryshe, P. G.** (2003). Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu Rev Genet* **37**, 485-511.
- Parenteau, J. and Wellinger, R. J.** (2002). Differential processing of leading- and lagging-strand ends at *Saccharomyces cerevisiae* telomeres revealed by the absence of Rad27p nuclease. *Genetics* **162**, 1583-94.
- Price, C. M.** (1997). Synthesis of the telomeric C-strand. A review. *Biochemistry (Mosc)* **62**, 1216-23.
- Qi, H. and Zakian, V. A.** (2000). The *Saccharomyces* telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein. *Genes Dev* **14**, 1777-88.
- Shore, D.** (2001). Telomeric chromatin: replicating and wrapping up chromosome ends. *Curr Opin Genet Dev* **11**, 189-98.
- Smogorzewska, A. and de Lange, T.** (2004). Regulation of telomerase by telomeric proteins. *Annu Rev Biochem* **73**, 177-208.
- Teixeira, M. T., Arneric, M., Sperisen, P. and Lingner, J.** (2004). Telomere length homeostasis is achieved via a switch between telomerase- extendible and -nonextendible states. *Cell* **117**, 323-35.
- van Steensel, B. and de Lange, T.** (1997). Control of telomere length by the human telomeric protein TRF1 [see comments]. *Nature* **385**, 740-3.
- van Steensel, B., Smogorzewska, A. and de Lange, T.** (1998). TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401-13.
- Vega, L. R., Mateyak, M. K. and Zakian, V. A.** (2003). Getting to the end: telomerase access in yeast and humans. *Nat Rev Mol Cell Biol* **4**, 948-59.
- Veldman, T., Etheridge, K. T. and Counter, C. M.** (2004). Loss of hPot1 Function Leads to Telomere Instability and a cut-like Phenotype. *Curr Biol* **14**, 2264-70.
- Watson, J. D.** (1972). Origin of concatemeric T7 DNA. *Nat New Biol* **239**, 197-201.
- Wellinger, R. J., Wolf, A. J. and Zakian, V. A.** (1993). *Saccharomyces* telomeres acquire single-strand TG1-3 tails late in S phase. *Cell* **72**, 51-60.

Table 1. Overview of Telomeric DNA Arrangements in Different Organisms

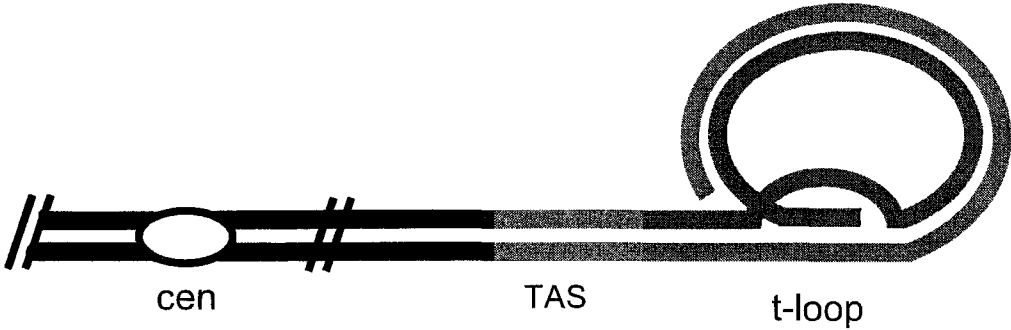
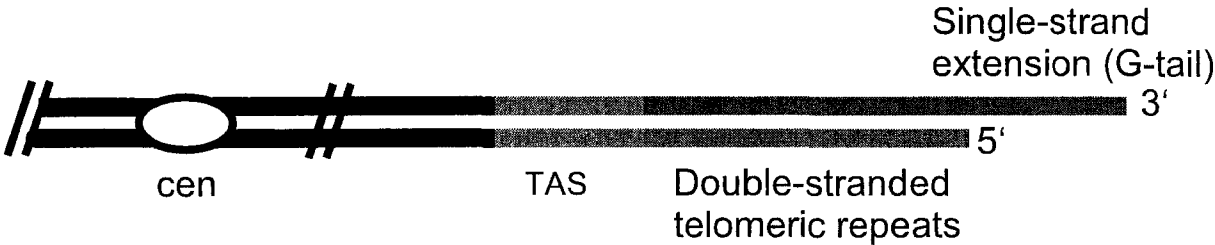
Organisms	Repeat sequence 5' → 3'	Length of ds repeats	Length of G-tail	Presence of t-loops
Protozoa				
Tetrahymena	T ₂ G ₄	250-300 bp	14-15 b	?
Oxytricha	T ₄ G ₄	20 bp (macronucleus) 3-20 kb (micronucleus)	16 b N/D	no yes
Euplotes	T ₄ G ₄	28 bp	14 b	?
Trypanosoma	T ₂ AG ₃	10-20 kb	~ 75-225 b	yes
Fungi				
Saccharomyces	(TG) ₁₋₆ TG ₂₋₃	~ 300 bp	12-14 b	?
Schizosaccharomyces	T ₁₋₂ ACA ₀₋₁ C ₀₋₁ G ₁₋₆	~ 300 bp	N/D	?
Plants				
Arabidopsis	T ₃ AG ₃	2-4.5 kb	~ 20-30 b	?*
Vertebrates				
Homo Sapiens	T ₂ AG ₃	2-20 kb	~ 75-225 b	yes
Many others	T ₂ AG ₃	2-100 kb	~ 75-225 b	?





? = unknown

N/D = not determined

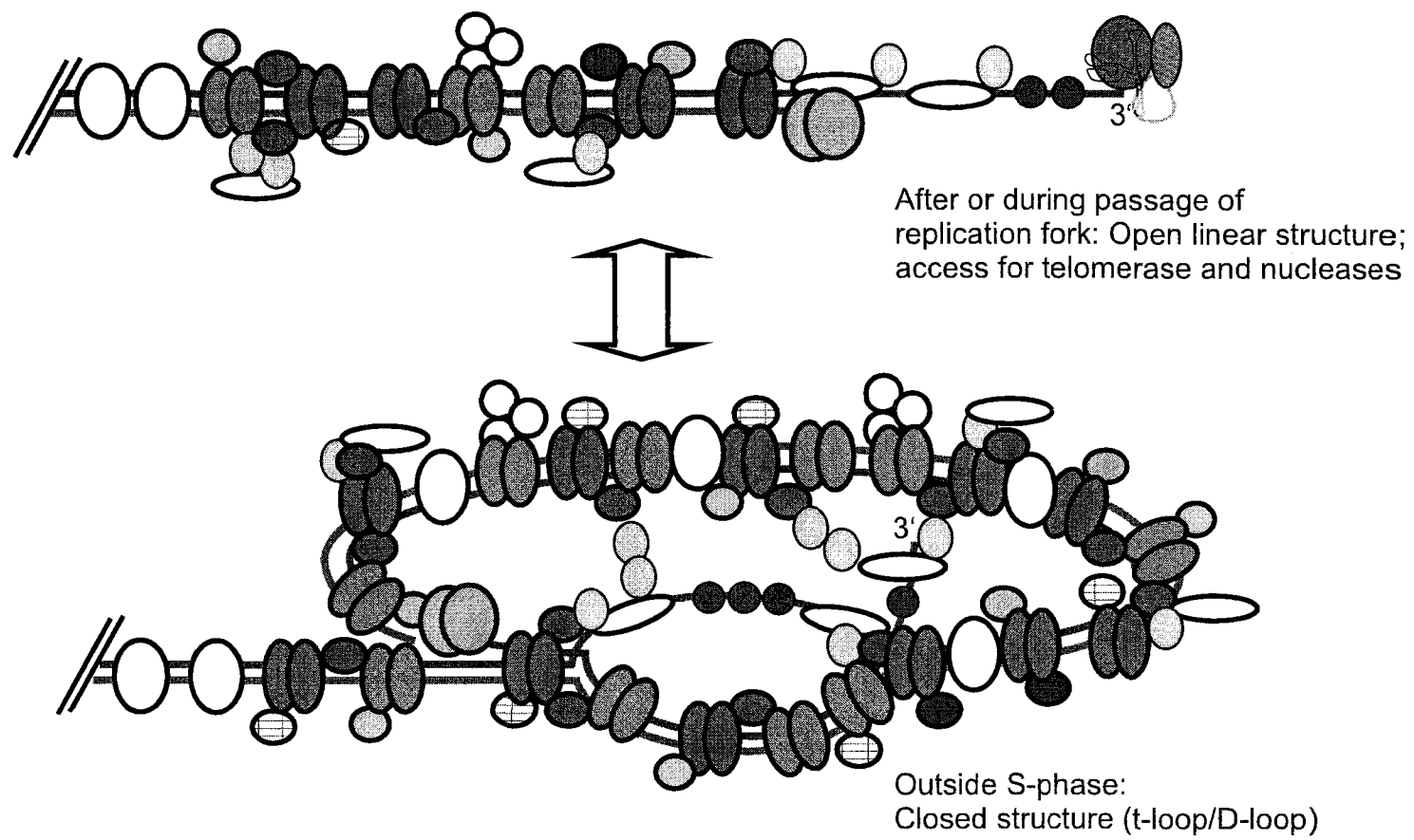
?* : There is evidence for t-loops in Peas.

DNA Configurations at Telomeres



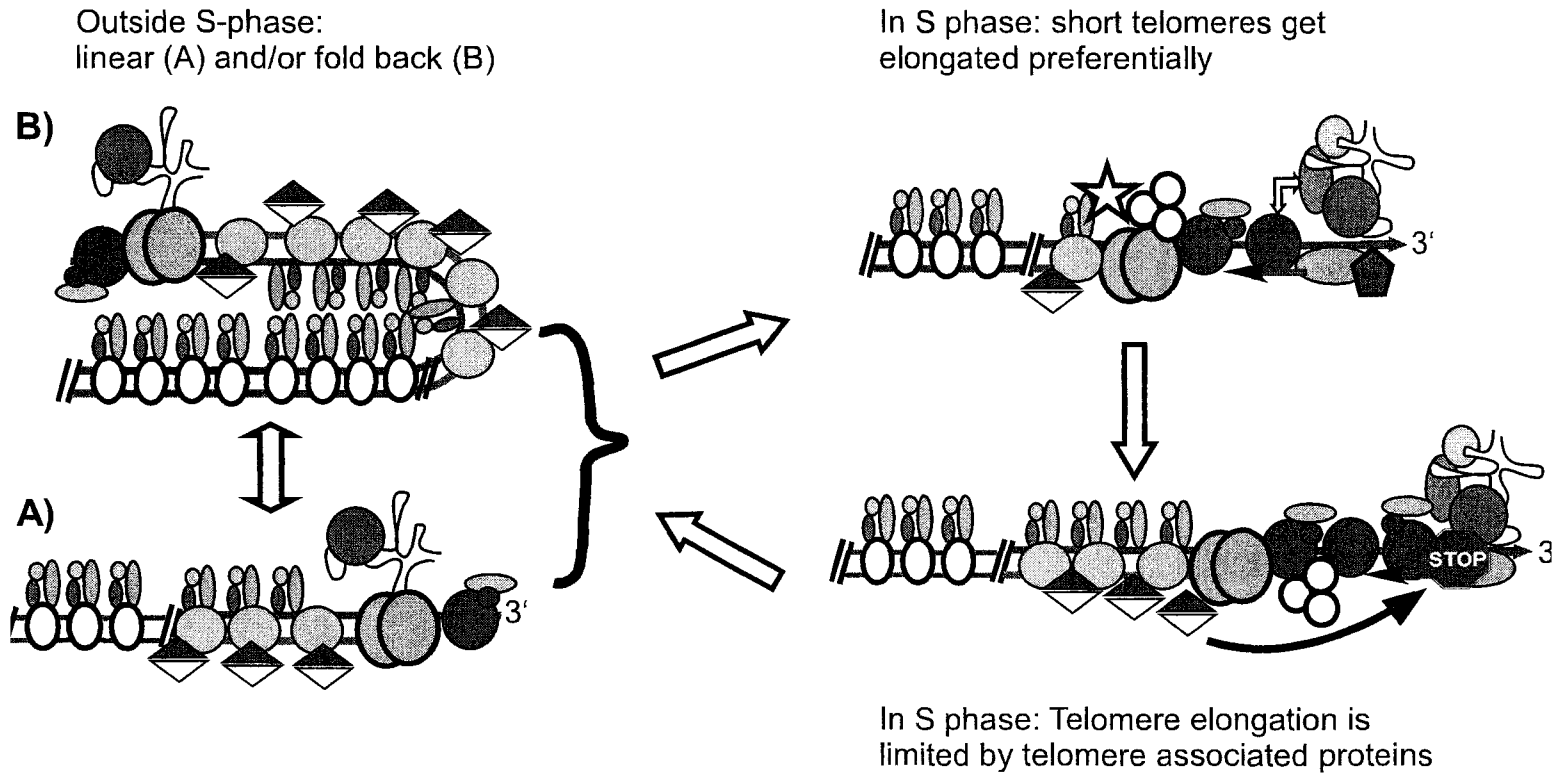
- | | |
|--|---|
|  G-rich telomeric DNA |  non-telomeric DNA |
|  C-rich telomeric DNA |  Telomere associated sequences (TAS) |

The cell cycle dependent changes of the structure and components of human telomeres.



Proteins directly associated with dsDNA	Proteins directly associated with extended G-strand	Other Telomere associated Proteins		Telomerase Components
		Major Components:		
TRF1	hPot1	hRap1	TIN2	hTERT
TRF2	hnRNPs (A, C, D)	PTOP/PIP1/TINT1		hTER RNA
nucleosome (H2A, H2B, H3, H4)		Also present:		DKC
		Ku	Tankyrase1	hEST1A/B
		BLM	Tankyrase2	
		PinX1	MRE11-complex	

The cell cycle dependent changes of the structure and components of yeast telomeres.



Proteins directly associated with dsDNA	Proteins directly associated with extended G-strand	Other Telomere associated Proteins	Telomerase components
nucleosome (H2A, H2B, H3, H4) Rap1p	Cdc13p lagging strand synthesis machinery (primase, DNA pol α)	Sir2/3/4 complex Ku Rif1/2p Ten1p/Stn1p MRE11-complex Tel1p Pif1p	Est2p TLC1 RNA Est3p Est1p